Pyranometer Accuracy in the PV Industry: Characterizing the Uncertainty in Performance Ratio Measurements

Master thesis submitted to Delft University of Technology in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

Management of Technology
Faculty of Technology, Policy and Management
Delft University of Technology

Student name: **Matthijs Ates**
Student number: **5077990**
Thesis defence date: **24 February 2026**

Graduation committee

dr. ir.	I. Bouwmans,	TPM E&I	Chair, First Supervisor
dr.	A.F. Correljé,	TPM E&I	Second Supervisor
dr. ir.	R.A. Verzijlbergh,	TPM ETI	Advisor
ir. C. J. van den Bos,		Hukx Sensor Technology	External advisor

Executive Summary

Solar photovoltaic (PV) power generation has seen an exponential growth over the past years and is expected to become the primary source of renewable energy by the end of this decade, according to the International Energy Agency (IEA, 2024a). The biggest subclass of this industry concerns megawatt to gigawatt scale projects. Due to the substantial financial investments involved in building and maintaining these 'utility-scale' power plants, accurate monitoring of their performance is of great interest to asset owners and other involved parties. Specialized PV plant performance metrics have been established to this end, that are used to monitor degradation rates over time, verify plant construction and maintenance service quality and to calculate the value of the asset. The most commonly applied metric is the performance ratio (PR) defined in the IEC 61724 standard.

PV performance metrics such as the PR rely upon measured data of the energy output and the solar irradiance incident on the power plant, which means that there is an inherent measurement uncertainty associated with these metrics. Previous studies have calculated uncertainty in the performance ratio to typically be between 2-8% (see e.g. Reise et al., 2018; Strobel et al., 2009). However, there is little discourse on the impact of uncertainty in decision-making based upon PR measurements, nor has there been an assessment on the effectiveness of accuracy improvements in the context of important use-cases of the PR. Furthermore, no previous calculations of PR uncertainty have explicitly included the contributions of individual uncertainty sources in the calculation

This study adds to the existing knowledge base by performing performance ratio uncertainty calculations with attention to the individual contributions of separate uncertainty sources, and by calculating the financial implications of this uncertainty in a relevant industrial application of the PR: payments associated with performance guarantees in Engineering, Procurement and Construction (EPC) and Operations & Maintenance (O&M) contracts. An investigation is made into the effect of solar irradiance measurement accuracy by comparing the uncertainty and associated financial risk for asset owners, EPC contractors and O&M contractors that can be achieved using standard high-accuracy pyranometers and industrial state-of-the-art pyranometers. The commercially available SR300-D1 pyranometer manufactured by Hukx was used to this end.

Performance ratio uncertainties were calculated to be between 2.8-3.9%, in line with published estimates. The SR300-D1 pyranometer was seen to outperform standard ISO 9060 class A specifications (which are the strictest standardized specifications) with a 0.6% lower uncertainty, constituting an accuracy improvement of 15%. By separately including individual uncertainty sources according to the GUM framework, assuming independence and maximal autocorrelation of uncertainty sources, it was calculated that the pyranometer directional response error is responsible for 35-60% of the total uncertainty, followed by PV module calibration uncertainty with 15-30%. ISO 9060 class A specifications further yielded large contributions by the temperature response and zero-offset A errors in pyranometers, while these were smaller and not substantially different from other errors in the SR300-D1 pyranometer.

Financial risk calculations were based on performance quality assessment as specified in EPC and O&M contracts, where the EPC or O&M contractor runs the risk of paying liquidated damages in case plant performance does not meet a contractual guarantee. For industry-practice performance tests, measurement uncertainty was seen to increase the financial risk of the contractor. The expectation value of the payable damages was most sensitive to uncertainty in cases where the performance of a plant is close to the guaranteed value. In this regime, the accuracy improvement of an SR300-D1 over standard class A specifications resulted in a 13% reduction of the EPC or O&M contractor's financial risk. In a hypothetical situation where the directional response error were to be halved, this reduction can increase to 35%. However, in case of severe PV system underperformance the expected damages were not sensitive to uncertainty due to statistical averaging effects. On the other hand, value-at-risk was seen to be linearly dependent on the measurement uncertainty even in this regime, and was significantly reduced by the SR300-D1 compared to standard ISO 9060 class A specifications. These results indicate that EPC and O&M contractors significantly benefit from higher-accuracy performance measurements. Furthermore, calculations were performed to quantify how different quality assessment testing types unfairly shift financial risk to either the contractor or the asset owner. Generally, measurement uncertainty was seen to increase financial risk of the EPC and O&M contractor, while commonly-encountered guard banding practices were seen to shift financial risk to the asset owner at an even greater degree. The calculation results can be used by asset owners and EPC or O&M contractors involved in performance tests to understand the implications and fairness of common contractual acceptance testing clauses.

This study is specifically focused on the role of pyranometer uncertainty in performance assessments, because this is the primary contributor (75-85%) to the total uncertainty budget. This research was done in collaboration with Hukx Sensor Technology, who are market-leading pyranometer manufacturers. Hukx provided expert insight into the workings and uncertainty sources of pyranometers, supporting proper inclusion in the uncertainty calculations. Hukx is recommended to reduce the directional response error for the SR300-D1, with the economic driver that this further reduces the financial risk associated with EPC and O&M contract conformity assessment.

The uncertainty calculations were preceded by an analysis of the utility-scale PV industry, the lifecycle of PV plants, important technical standards, the applications of PV performance metrics and the contractual technicalities and variations of contractual performance guarantees for construction and maintenance services, establishing the contextual foundation for this work. Particular attention was dedicated to understanding how contractual performance guarantees are structured in practice. This sector analysis is based on a literature study and a small set of interviews with industry professionals that served to verify and expand upon what was compiled from literature.

Certain aspects of this work are limited in scope and merit further investigation. The uncertainty calculations rely upon assumptions of independence and autocorrelation, and can be improved upon by including covariances between uncertainty that are more physically accurate. This can help identify further accuracy improvement avenues beyond the directional response error identified in this work. Uncertainty due to soiling is also excluded. The code developed for this work is freely available and

can be used to perform a comprehensive study on soiling uncertainty in different climates that can help inform cleaning schedules. Finally, the sector analysis consisted of a limited qualitative component and was largely based on secondary literature. The applicability of the financial risk calculations to practical situations can be extended by a more thorough investigation of industry performance guarantee practices.

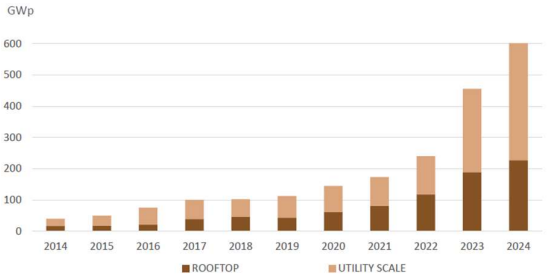
This study has shed light on how impactful irradiance uncertainty is in performance ratio calculations and to financial interests of asset owners, EPC contractors and O&M contractors in plant acceptance testing. The results demonstrate the value of minimizing uncertainty and the importance of its consideration in deciding upon contractual performance guarantees. Through the novel approach to include uncertainty sources individually, promising insights for further pyranometer improvement were obtained. The financial calculations can be used by Hukx to underpin the economic relevance of high-accuracy pyranometers for clients involved in projects with performance guarantees (such as turnkey projects), and can support EPC and O&M service contractors and asset owners in assessing the fairness of and risks associated with performance guarantee acceptance testing. Hence, this study provided actionable insights for pyranometer manufacturers and supports better-informed decision making with respect to performance guarantees in the utility-scale PV sector.

Table of Contents

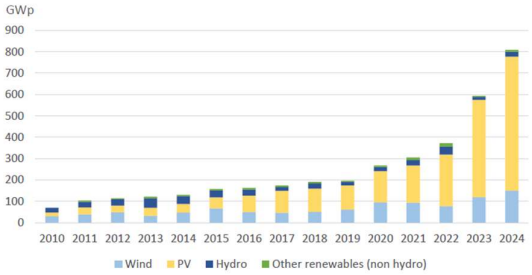
1. Introduction

The worldwide shift to sustainable energy has led solar photovoltaic (PV) power generation to see a dramatic rise as an energy source in the past 20 years. As of 2024, total installed solar PV capacity is estimated to be around 2.25 TW (Masson et al., 2025), accounting for over 5% of worldwide energy generation. Total installed solar PV capacity is expected to triple by 2030 and become the largest source of renewable energy, overtaking wind and hydro generated power according to the International Energy Agency (IEA, 2024a), with 600 GW of new capacity installed in 2024.

Solar PV is special as an energy source because it has allowed homeowners, traditionally on the demand side of the electricity market, to start generating their own electricity. This 'rooftop' power generation accounted for one-third of newly installed capacity in 2024. However, two-thirds of this newly installed capacity falls into the so-called *utility-scale* category, which refers to large, grid-connected solar power plants. The utility-scale PV sector will be the focus of this work.



1.1a: Amount of newly installed solar PV power generation capacity each year, split between rooftop and utility-scale capacity.



1.1b: Amount of newly installed power generation capacity from renewable sources each year, split between different renewable sources.

Figure 1.1: Growth of the solar PV industry over the past decade. Figures taken from Masson et al. (2025).

Figure 1.1 shows the exponential growth solar PV as an energy source over the past years. With its growing prominence in the global electricity market, solar PV power generation is starting to evolve into a mature industrial activity. The IEA estimates that a total of 500B USD has been invested in installing new capacity in 2024, more than the combined annual investment into all other renewable energy sources (IEA, 2024b).

With the rise of utility-scale PV as a global energy source and the commensurate increase of capital investment, accurate monitoring of performance and energy output of PV plants becomes increasingly important. Solar PV is inherently unpredictable as a power source due to its dependence on meteorological conditions, and the performance assessment of solar power plants is dependent on an even broader variety of factors. Standards have been developed to codify the performance assessment of utility-scale solar PV power plants, most prominently the IEC 61724 set of standards and the ASTM E2848 standard (International Electrotechnical Commission, 2021; American Society for Testing and Materials, 2023). Because of the sudden rapid growth of this industry, standards and best-practices are still developing, with the IEC 61724 standard being under review as of the writing of this thesis.

One of the most relevant metrics to determining PV system performance is the performance ratio (PR). The PR is defined in the IEC 61724-1 standard as the total energy output divided by the theoretical benchmark energy output of the solar power plant given the measured in-plane solar irradiance, over a specified time. It normalizes the achieved energy yield as a percentage of the nameplate rated yield under the measured environmental conditions.

Other performance metrics exist as well, such as the energy performance index defined by the IEC 61724-3 standard and the linear regression model outlined in the ASTM E2848 standard, which both compare realized output to the predicted power output from a computational model of the plant.

Metrics such as the PR are designed to provide a quantification of the overall performance of the power plant while accounting for the uncontrollable nature of the input resource (solar energy), and have multiple uses in industry. They are applied to monitor PV asset performance and degradation over time, information that in turn informs operational decisions such as retrofitting or end-of-life commissioning, and they are used to determine the financial value of the asset, which is of great significance when selling the asset or during corporate mergers and acquisitions.

Of particular interest in this work is the use of performance metrics in service quality guarantees in Engineering, Procurement & Construction (EPC) and Operations & Maintenance (O&M) contracts. As a quality assurance method, external EPC or O&M service contractors provide a guaranteed value of the PR, or another performance metric, that the plant will achieve during a contractually agreed upon performance test. These guaranteed values are enshrined in EPC and O&M contracts and underpinned by clauses specifying reimbursement of the asset owner in case they are not achieved during the test.

The PR and other PV plant performance metrics are calculated from power output, solar irradiance and environmental data such as the temperature. These quantities are determined through measurement and are hence inherently uncertain, which translates to uncertainty in the performance metrics that are calculated from this data. Studies have quantified the uncertainty in the performance ratio to be between 2-7% (Strobel et al., 2009; Basson & Pretorius, 2016; Reise et al., 2018; Özkalay et al., 2022).

Figure 1.2: Desert Sunlight Solar Farm, a 550 MW capacity power plant in California, USA. Estimated capital cost of this facility exceeds 800 million USD based on a capex of 1.60 USD/W. Image source: Fortune, 2016.

Calculations and decisions based on the performance ratio are therefore significantly influenced by uncertainty, while the power plants they concern cost millions to hundreds of millions USD in capital to build and maintain. Estimations calculated in this work show that 1% underperformance can translate to over 1 million USD in discounted foregone revenue over the operational lifetime of the powerplant. However, despite the well-documented magnitude of the uncertainty in performance metrics, little attention has been given to the implications of this uncertainty with respect to the financial interests connected to performance ratio measurements.

A case where this is especially relevant is during the conformity assessment to the previously mentioned performance guarantees defined in EPC and O&M contracts. These assessments involve comparisons between the guaranteed PR and an uncertain measured PR, with potential liquidated damages (in the range of millions of USD) determined by the value of this measured PR. Specifics of these tests can vary on a case-to-case basis. However, no investigations have been done into how measurement uncertainty affects various types of contractual agreements and their assessment accuracy, what the financial risk due to uncertain PR measurements is for the parties to this test and whether this risk can be reduced through accuracy improvements.

Furthermore, previous studies quantifying the uncertainty in performance metrics, including the cited publications on the PR, are based upon flat-percentage uncertainty estimates for pyranometer measurements, taken to be between 2-5%. While pyranometer measurement uncertainty for daily irradiance indeed falls within these estimates (Hukx, n.d.-a; Konings & Habte, 2015), the uncertainty of a pyranometer is in reality composed of many different sources contributing to the total uncertainty. These contributions are not all identical either, some sources contribute to the total more significantly than others. At present, no studies have done an assessment on how these individual pyranometer uncertainty sources contribute to the total uncertainty in the performance ratio. A calculation that keeps individual uncertainty sources explicit has multiple advantages. It can for instance be used to compare the achievable uncertainty in PR using pyranometers of different specifications, and to inform the most promising avenues of improvement in current state-of-the-art pyranometers.

This study aims to add to the existing literature in two ways. First, by performing methodological uncertainty calculations that keep explicit the different internal pyranometer uncertainty sources, with the purpose to obtain insight into the most significant contributors. Second, by analyzing the current performance guarantee conformity assessment practices in industry in the context of financial risk for the involved parties due to measurement uncertainty. The potential improvements identified and quantified through the uncertainty calculations can then be placed in this practical context. Together, this will provide insight in the financial uncertainty in acceptance testing that allows for better-informed decisions to be made in tests to performance guarantees, and to quantify the importance of measurement uncertainty to the interests of stakeholders in these tests: asset owners, EPC contractors and O&M contractors.

The literature gap that this work attempts to bridge is in a multidisciplinary intersection of utility-scale PV legal practices, state-of-the-art irradiance measurement technology, PV system performance assessment standards and the mathematics of uncertainty propagation and financial risk assessment. Therefore, it is important to establish an understanding of the utility-scale PV industry: its common practices, relevant standards and important stakeholders, with special attention for the importance of irradiance measurements to the industry. This sector analysis is the theoretical basis for many aspects of this work and allows the research topics to be placed in the context of the utility-PV industry at large as we proceed to bridge the gap.

The sector analysis is in large part based upon a review of published literature on these matters. Specific attention will be given to the role and structure of performance metrics and performance guarantees in industry. This analysis will be accompanied by a small set of interviews with experts from the solar PV industry, that serve to verify and expand upon the findings of the literature study.

The uncertainty calculations of the PR are performed following the methodology prescribed by the Guide to the Expression of Uncertainty in Measurement by the Joint Committee for Guides in Metrology (JCGM, 2008). Some assumptions and approximations are made with respect to the calculation methodology and uncertainty characteristics, which will be discussed in further detail in chapter 3.

Uncertainty sources affecting PR measurements will have uncertainty magnitudes taken from relevant standardized specifications where possible, and estimated from literature otherwise. Notably, the specifications for pyranometer uncertainty will be taken as the ISO 9060 class A specification, which is the most accurate class defined in this authoritative standard for pyranometer measurement specifications. A comparison will be performed against the specifications of a state-of-the-art commercially available pyranometer, the Hukx SR300-D1. This work has been performed in collaboration with Hukx Sensor Technology, who are market leaders in solar irradiance measurement equipment. Their technical expertise supported proper implementation of relevant uncertainty sources and helped interpret the results of the uncertainty calculations.

The results of the uncertainty calculations will be placed in the context of conformity assessment to, and damages associated with, performance guarantees. This includes calculations of the probability of passing various acceptance testing clauses identified from literature, the expected payable damages and value-at-risk for the EPC and O&M contractors in these tests,

and the expected errors in the owed damages. The calculations will also explicitly quantify financial risk reduction for EPC and O&M contractors resulting from higher-accuracy solar irradiance measurements.

The results will be combined and analyzed in a discussion, which will also restate and reflect upon the important methodological assumptions, comment on the limitations on the applicability of this study to the greater utility-PV sector and the potential avenues for future research.

The research question and subquestions that we aim to answer through this methodology are stated as follows:

RQ: *What is the influence of individual uncertainty sources in state-of-the-art pyranometers to the uncertainty of utility-scale solar PV plant performance metrics, specifically the performance ratio, and what are the most impactful avenues for pyranometer improvement?*

S1: *Where are solar irradiance measurements of importance in the lifecycle of a utility-scale solar PV power plant?*

S2: *Which metrics are used to determine the performance of solar PV power plants, and for which purposes are performance metrics used in industry?*

S3: Which sources of uncertainty constitute the total uncertainty in the performance ratio, what is their relative weight and what are the most promising areas of improvement?

S4: How is uncertainty in performance metrics included in decision-making by industry?

The findings of the literature review are presented in chapter 2, which includes an outline of the lifecycle of utility-scale solar PV power plants, compiled findings on important contractual arrangements and performance assessment methods, measurements and standards. Important foundational definitions on the interpretation of measurement uncertainty are given, and the current role and knowledge gaps with respect to measurement uncertainty in the solar PV industry are discussed.

The research methodology is outlined in chapter 3, which includes an outline of the sector analysis methodology and interview strategy, the uncertainty calculation methodology and the mathematical basis of the financial calculations performed in this work. The results for the sector analysis and the uncertainty and risk calculations are presented in chapter 4.

A discussion of this work is given in chapter 5, that includes an interpretation of the methodology and results, an assessment of limitations of the applicability of this work and recommendations for future research. A conclusion is given in chapter 6.

Three appendices are attached to this work. Notably, appendix C provides a link to the GitHub repository containing the python code used for the uncertainty calculations, with an accompanying user manual.

2. Theoretical Background

2.1 Utility-Scale Solar PV Power Plants

Solar PV power generation has the uncommon characteristic in the fact that it can be generated cost-effectively from the scale of a couple watts, for instance using an integrated panel in a watch or calculator, to the scale of gigawatt power plants. In this work, we will focus on utility-scale solar powerplants. The term “utility-scale” has no strict definition, but is rather an umbrella term used in industry to refer to large plants. Various different definitions are employed in literature, with the term usually referring to plants with an installed capacity greater than 1-5 MW (NREL, 2018; Seel et al., 2024; Wolfe, 2013). In some cases, no strict definition is given at all. In this research, we simply use it to refer to megawatt-scale ground-mounted solar PV powerplants.

2.1.1 The Solar Power Plant Lifecycle

The development cycle of solar PV powerplants is in many ways similar to other large energy infrastructure projects. Usually, there is a company or consortium looking to build a solar power plant that drives the early development stage. This party performs site prospecting, acquires relevant permits, designs the power plant, signs agreements with relevant contractors and secures financing for the project. After the development stage, the plant is built by the engineering, procurement and construction (EPC) contractor. Once construction is finished and the build-quality of the plant is verified, the plant enters into regular operations under the portfolio of the final asset owner. During operations, the asset owner usually outsources plant operations and maintenance (O&M) to a dedicated contractor. Utility-scale solar PV powerplants have a regular operational lifetime of around 20-30 years (Curtis et al., 2021), although this is dependent on environmental conditions and maintenance quality.

We will divide the powerplant lifecycle in three phases, the pre-construction, construction and post-construction/operational phase. We will briefly overview these phases and the role of solar irradiance measurements therein. This section is not a detailed review of solar power plant construction, operations, assessment and finance. For a comprehensive discussion of these topics we refer to a publication from the International Finance Corporation (2015). A visualization of the lifecycle model outlined in this section is provided in figure 2.1.

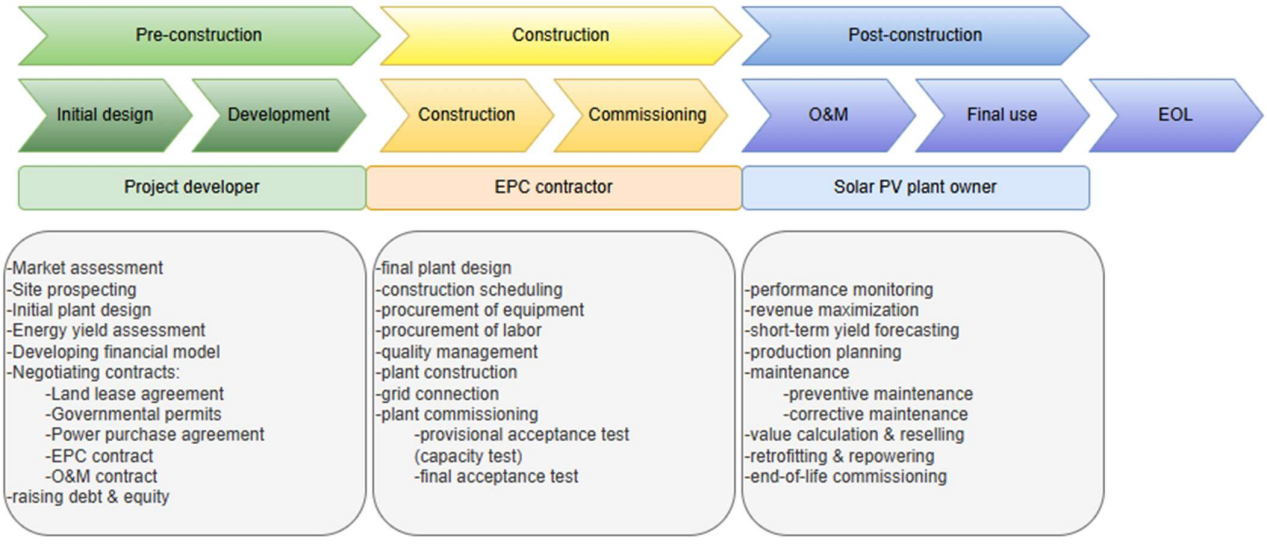


Figure 2.1: Diagrammatic representation of the solar PV power plant lifecycle. The upper bars denote the phase of the lifecycle, the secondary bars the important activities during the phase, the third bar denotes the main driver of the activities during the phase, and the grey blocks list important activities conducted during each phase.

During the pre-construction phase the project developer is the main driver. The developer aims to build a business case which is deemed 'bankable' by investors. In general terms, this means that the necessary permits should be obtained, contracts with the land owner, grid operator, power off-taker, EPC and O&M contractors should be negotiated, plant design and implementation plan should be finalized and a bankable financial model should be prepared by the project developer. This comprehensive documentation is necessary for raising the required debt and equity from investors and money lenders during the main financing round, as financers require detailed reports on projected costs and cash flows for their risk analysis on the prospective investment.

It is common practice to include an energy yield assessment with the business case. An energy yield assessment is the usage of weather data over an extended period to calculate an estimated energy yield of the prospective power plant. Usually, the weather data comprises a historical weather satellite dataset and data from a ground measurement campaign. The ground measurements typically last a full year, to capture the entire range of seasonal conditions and is combined with historical satellite data (for instance from the preceding 20 years) to 'calibrate' or adapt the satellite data to the site under consideration (Cebecauer and Suri, 2016; Polo et al, 2016). This calibrated dataset is then used to extrapolate expected energy yield over the expected plant lifetime. Combining satellite and ground data helps to average out meteorological anomalies that may occur during the field campaign, while helping to calibrate the usually less accurate satellite data to the site conditions. The extrapolated irradiance forecast is used as input for the computational model of the powerplant to help refine power plant design and to obtain energy yield projections.

Once design, contracts and financing are finalized, the construction of the plant commences. During the construction phase, the chosen EPC contractor is the main driver of the project and is assigned the task to carry out detailed engineering, component and labor acquisition, and oversee construction and maintenance. Additionally, the EPC contractor assumes the responsibility to deliver the power plant at a fixed budget, before a fixed deadline and with a performance guarantee for the completed power plant (PWC, 2024). After construction is completed a there is a transitionary period during which the performance of the power plant is verified. We will discuss the technicalities of performance assessment in-depth in section 2.2, but usually this performance testing is two-step process. First, upon completion of the project a capacity test lasting a few days is performed, and second an energy yield test is performed lasting a couple months to a few years.

Only after the plant passes both of these tests is the final payment to the EPC contractor is released, although operational responsibility of the powerplant is usually transferred to the owner after the first test is passed. Who this owner is depends on the exact financing structure of the project. In any case, operations and maintenance (O&M) of the plant will usually be performed by a dedicated O&M contractor. The O&M contractor is responsible for continuously monitoring plant performance and ensuring efficient operation. Generated energy will be bought by the power off-taker according to the agreed upon terms in the power purchase agreement (PPA). Meanwhile, the asset owner is responsible for the greater strategic management decisions related to revenue maximization, retrofitting and end-of-life decommissioning. End-of-life of solar power plants is usually reached after a period of 20 to 30 years (Curtis et al., 2021), although this can be dependent on environmental conditions.

The above is but a brief overview of the complicated process of utility-scale solar PV development, construction and management. We will treat specific sections of this lifecycle in more detail in the coming sections, but for now let us make some general comments on this model.

First of all, the three phases distinguished here do not have strict boundaries. Realistically, engineering work and procurement of components, as well as preliminary construction work, may already start before the development phase is truly "finished". Similarly, regular plant operations may commence before mechanical construction of the full plant is finished (when sections are connected to the grid while others are already being built). In addition, the plant enters regular operations prior to passing the final acceptance test, which is primarily a legal process.

Secondly, there is not always a strict division of responsibilities in this process. Many companies are vertically integrated, assuming some or all responsibilities related to development, financing, EPC, O&M and asset ownership. This integration can have implications on how certain elements of the development process take shape. For instance, if the EPC contractor is also the final owner, acceptance testing may be less stringent compared to turnkey projects (where the EPC contractor and final asset owner are two different entities). Horizontal integration is also possible, e.g. in the case where a utility company develops a solar PV plant in-house, in which case a PPA may not be applicable.

Moreover, not all relevant parties involved in the development process are included here. Some specific activities, such as energy yield assessment, may be outsourced to dedicated third-parties which are not mentioned in this section. A complete supporting ecosystem of specialized service providers exists that may, or may not be, involved during various stages of the sketched lifecycle model.

The above examples aim to illustrate the vast amount of variations of the lifecycle that realistically exist. The rudimentary model presented here is meant to provide context to common practice in the utility solar PV lifecycle which are of interest to this study, specifically moments where solar irradiance measurements are of importance.

2.1.2 Common Contractual Arrangements in Utility Solar PV

Now that an outline of the solar PV lifecycle has been established, let us elaborate on some important common contractual arrangements that are entered into during solar PV development as they are described in literature. Relevant to this research are the EPC and O&M contracts, and the PPA. This section will not include all aspects of these contracts but will instead treat specific parts relevant to this thesis. Full reference readings are provided for each contract type.

Engineering, Procurement and Construction Contracts

The EPC contract codifies the obligations of the EPC contractor and the payment regulations for the delivered services. It contains component such as timelines, performance guarantees, payment structure and legal contingency clauses. For extended reading we refer to the publications on which this section is based, by the IFC (2015), Stoel Rives LLP (2022), PWC (2024) and SolarPower Europe (2021).

Usually, the power plant design is already decided up to a certain level of detail before the EPC contractor takes control of the project. Elements like array positions, tilt angles, inverter substation locations and grid connection points may already be fixed, and potentially there are specifications on the type or brand of components, such as modules or measurement equipment, that may be installed. However, there often remain some degrees of freedom along which the EPC contractor can make decisions and try to maximize its profit margins.

An important part of EPC contracts is the payment structure. We will refer to the total payment made to the EPC contractor over the contract duration as the ‘contract price’ or ‘contract value’. After an advance payment to cover initial costs for the EPC contractor, subsequent payments are released upon achieving specific milestones, such as when certain sections of the power plant are connected to the grid. Once all construction work is finished the project will have reached mechanical completion and a provisional acceptance test is performed. Depending on contract specifics, a payment may be tied to passing this test. However, the final payment to the EPC contractor is only released after the plant has passed the final acceptance test, which may take as long as 2 years and can therefore also be viewed as a ‘warranty period’.

According to literature, the payment tied to passing this final test can be as high as 5-15% of the total contract value (PWC, 2024; SolarPower Europe, 2021). However, in case a project fails to pass the final acceptance test this does not mean the final payment is withheld by definition. Instead, there are clauses in place that define the liquidated damages that are owed by the EPC contractor to the asset owner, usually calculated to cover the net present value of the foregone energy yield over the plant lifetime due to the lower performance (IFC, 2015; SolarPower Europe, 2021). Moreover, in cases of severe underperformance the asset owner can retain the right to reject the constructed power plant in its entirety, and recover all

project costs on the EPC contractor. The specifics of these processes are contractually dependent and may differ on a case-to-case basis.

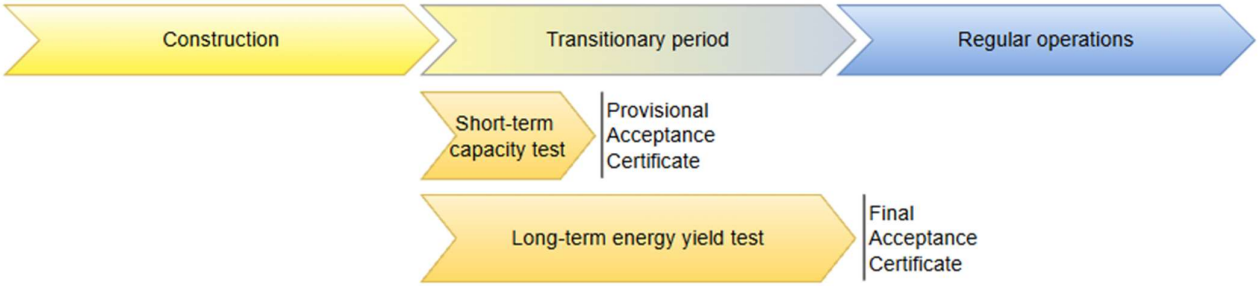


Figure 2.2: Schematic illustration on the transition process between the construction and post-construction phase. The transitionary period commences once mechanical construction is completed. This representation is illustrative purposes; variations on this model exist.

Although a full treatment of solar PV plant performance assessment is postponed to section 2.2, the topic is strongly tied to the acceptance testing process. In general terms, however, the parties entering into the EPC contract agree upon specific performance metrics to be used in the testing process as well as minimal values the measured metrics must meet for the plant to pass the tests. The power output and various environmental factors required for calculating the agreed performance metric are then measured over an agreed amount of time, and the achieved performance is compared to the contractually agreed upon performance. If the plant does not pass the test, the clauses outlining damages owed to the asset owner will be triggered. Otherwise, the contractually agreed upon payment to the EPC contractor will be released.

As stated, there is usually a initial acceptance test lasting a few days (after which a provisional acceptance certificate, or PAC, is issued) and a general acceptance tests lasting 1-2 years (to which the final acceptance certificate, of FAC, is tied). The initial acceptance test usually tests whether the plant meets capacity expectations and is therefore performed on days where there is sufficient solar irradiance. The final acceptance test includes all days over an extended period, such that it covers a full range of meteorological conditions.

Not all utility-scale PV projects include ECP contracts, acceptance tests and damage payments. Some companies are vertically integrated along the value chain and perform EPC work in-house, in which case acceptance testing is an internal quality assurance process. The contracts outlined in this section are a staple of turnkey development projects, where the EPC work is performed by a dedicated external contractor. Literature indicates that this is the most common type of utility-scale PV development project, although no exact quantifications of this claim are provided (IFC, 2015; PWC, 2024).

Operations and Maintenance Contracts

This section is based on publications by the IFC (2015), National Renewable Energy Laboratory (NREL, 2018), the International Energy Agency Photovoltaic Power Systems Programme (IEA PVPS,2022), SolarPower Europe (2025) and the Solar Bankability Project (2016).

The O&M contract specifies the scope of services provided by the O&M contractor and the indicators by which maintenance quality is defined. O&M service contractors handle the monitoring of data streams, coordination with the grid operator, preventive and corrective maintenance, plant performance optimization, regular reporting to the asset owner and other tasks specific to local environmental conditions. Preventive maintenance concerns tasks such as inspections and cleaning, corrective maintenance concerns fault diagnosis and (emergency) repairs.

During operations there are various different KPIs to monitor plant performance and maintenance quality. A comprehensive overview of these is provided in the publication by SolarPower Europe (2025). Common KPI's which are tracked include dedicated PV system performance metrics (to be discussed in section 2.2.1), plant availability (operational uptime of the plant compared to total available operating time) and response time (the time between a fault occurring and the arrival of a

technician). According to literature, service quality is often contractually specified through guaranteed values of these KPIs that the service provider will achieve, such as a minimal plant availability or a minimally achieved performance ratio (see section 2.2.1). The KPI's are evaluated over regular intervals, for instance on a yearly basis.

O&M contracts also contain clauses determining the damages to the plant owner in case the service provider falls short on a specific guarantee. Literature mentions two ways of determining the amount of owed liquidated damages, either setting it equal to the value of the missed energy yield due to plant underperformance or as a fixed payable amount per percentage-point underperformance with respect to the relevant metric. Note that these clauses bear great similarity to how liquidated damages in EPC contracts are defined.

It is stated in the aforementioned literature that the specific service quality KPI's tend to differ on a contract-to-contract basis. For instance, guarantees with respect to PV plant performance metrics such as the performance ratio tend to meet resistance from third-party O&M providers, since plant performance is in large part determined by the construction quality of the plant. In case the O&M provider is not the same company as, or affiliated with, the EPC contractor, they are typically unwilling to assume responsibility in this regard.

The publication by SolarPower Europe (2025) explicitly states a shift away from plant performance guarantees in the third-party O&M market, in favor of O&M-specific KPI's such as plant availability. The publication mentions that instead contracts warrant an investigation if the plant performance decreases below a certain threshold, without immediate financial consequences for the O&M contractor. According to NREL (2018) and SolarPower Europe (2021), outsourcing O&M activities to dedicated third-parties is common industry practice, since these can deliver cheaper services by leveraging economies of scale.

Power Purchase Agreements

The power purchase agreement is a contract between the owner of the powerplant and the grid operator or utility company purchasing the generated electricity. PPAs specify the amount of power to be delivered to the grid and the price at which this power is purchased, among other clauses. They are important to project developers since they provide a basis for the projected incoming cash flows of a project, thereby reducing the financing risk for investors.

Multiple different types of PPAs exist, see for instance Mittler et al. (2025) and the report by Baringa Partners LLP (2022), but a broad distinction can be made between *physical* PPAs and *financial* or *virtual* PPAs.

In a physical PPA, the asset owner and utility company agree upon an amount of energy that must be delivered to the grid and a fixed tariff to be paid per delivered kWh or MWh of energy. For energy sources reliant on unpredictable external factors such as solar PV, the agreed upon amount of energy to be delivered is typically defined as a range in which the total annual energy delivery must fall. The agreement also contains clauses when the total delivered energy falls outside this specified range. In case of overproduction, the offtaker usually retains the option to purchase excess produced energy at a specific tariff, which may or may not differ from the regular tariff. Only if the offtaker elects to forego this option is the asset owner allowed to sell the excess production on the market. In case of a shortfall in energy yield, the offtaker is owed reimbursement for the energy that was not delivered equal to the yield deficit times an agreed upon tariff.

The electricity pricing of physical PPAs is typically subject to an escalation clause to account for inflation. Assuming an annual inflation rate of 2% and a contract term of 25 years, the net present value of a fixed tariff in the last year of the contract is discounted to approximately 60% of its value in the first year. This is counterbalanced by applying an annual escalation rate to the agreed-upon tariff. This escalator can be a fixed percentage specified in the contract or coupled to an inflation or price index.

In a financial PPA, the asset owner sells the produced energy directly to buyers on the electricity market. These types of PPA are sometimes also referred to as virtual PPAs, because even though the electricity is sold to a specific buyer, the energy is

still delivered to the electricity grid; no physical delivery of electricity to the buyer takes place. Financial PPAs allow asset owners to sell their generated electricity against market price, which gives them agency to try to increase their revenue with a trading strategy. This works both ways, however, and can also result in a loss compared to a fixed price physical PPA.

An important aspect in PPAs is the handling of *curtailment*, which is the deliberate reduction of energy production below maximally achievable production. Modern energy grids notoriously suffer from congestion issues during peak hours. Grid operators therefore may instruct asset owners to limit, or *curtail*, their energy production. However, curtailment can also be caused by the offtaker's bidding strategy on the energy markets. Naturally, the seller wants to be compensated for curtailed production. Often, a distinction is made between curtailment of voluntary and involuntary nature; curtailment due to offtaker market strategy will have to be compensated, while curtailments due to grid congestion are allocated to the seller (Stoel Rives LLP, 2022). Curtailment handling is understandably a topic of difficulty in PPA negotiations, thus the exact terms can differ on a case by case basis.

Beyond pricing and annual production specifications, PPAs may require the seller to provide the grid operator a detailed intra-day and day-ahead energy production forecast. These forecasts help the grid operator to facilitate stable grid operation conditions. Deviation from this forecast may result in fines to be imposed on the asset owner.

As a final remark, PPAs signed during the development phase often specify a commencement date at which the seller will need to start delivering energy to the grid, and contain clauses for maximal installation downtime, and the asset owner may incur fines in case these requirements are not met. Thus, while providing certainty with regards to projected cash flow and hence debt service coverage to investors, there is added risk to the developer in case construction fails to complete on time. The developer may in turn try to shift this risk to the EPC or O&M contractors in the respective contracts.

2.1.3 Technical Background of Solar PV

In this section we will discuss the main components of solar PV power plants, major variations that exist in plant design, factors impacting power generation and recent technological developments in the utility-scale solar PV industry.

A utility-scale solar PV power plant usually consists of many rows of metal racks on which solar PV modules are mounted. The modules on these racks are angled to catch the maximal amount of solar irradiation as possible. The angle of the modules with respect to the surface is called the *tilt angle*. Sets of modules are connected in series to form *strings*, strings are connected in parallel to form *arrays*. The plane in which all modules of an array are angled is also called the *plane of array*. Solar PV modules deliver direct current, which is converted to alternating current by *inverters*, this is usually done at the level of strings or arrays. The generated power is delivered to the grid as alternating current.

Solar PV modules consist of a multitude of *solar cells*, which convert solar radiation to electrical energy. A solar cell is a semiconducting material doped to become a pn-junction, where the excitation of electrons (i.e. creation of electron-hole pairs) caused by light absorption causes a charge splitting between the p- and n-doped regions. The separation of charge carriers causes a potential to build inside the solar cell, which causes a current to flow from one side to the other when the two sides are short-circuited. This current can be used to perform useful work by connecting a load to this circuit. The characteristics of a solar cell are in large part defined by the used semiconductor and doping levels. Modern solar cells use highly specific semiconductors to maximize efficiency or to optimize other qualities of the module.

Traditionally, solar PV modules were directional, with only one side being able to catch incident light and convert it to energy. However, modern technology has enabled specific modules to also catch and convert irradiance incident on the back of the solar module. This has further increased the efficiency of modern solar PV modules. However, not all PV systems employ such modules, since they can be more expensive and the additional yield from the back of the module may be deemed insignificant.

Another method which has been developed to maximize yield is by placing PV modules on tracker systems that ensure the PV modules are always oriented to capture as much irradiation as possible. However, solar tracking may lead to shading of PV modules by modules in the next array. To mitigate this, arrays may need to be placed further away from each other, thus increasing the amount of land required for the plant.

The efficiency at which solar PV modules convert incident light to electrical energy is dependent on several factors. Prominent among these are the module's *spectral sensitivity* and its *temperature*. The efficiency at which a PV module can convert an incident photon to electrical energy is dependent on the wavelength of this photon. The function of the conversion efficiency with respect to photon wavelength is called the module's spectral responsivity and is largely determined by the specific semiconductor used in the solar cells. Module temperature negatively influences the PV module's efficiency over the entire spectrum. The sensitivity to temperature is again determined by the physics of the semiconductor used in the cells. Notably, PV modules intrinsically heat up when converting light to electrical energy. The temperature, and thus efficiency, is thus determined by PV cell properties, ambient temperature, available irradiance and the wind speed (wind helps modules lose thermal energy by facilitating convection).

2.1.4 The Uses of Solar Irradiance Measurements

The applications of solar irradiance measurements in the solar PV industry can be broadly divided in two categories: to calculate plant performance metrics and to inform forecasting models that calculate future irradiation. Plant performance metrics and irradiance forecasts are subsequently used for a great variety of purposes at different stages of the plant lifecycle, which will be summarized in this section. This section will not treat the inclusion of uncertainty, this is postponed to section 2.4.1.

Performance Metrics: Operational Decisions and Quality Guarantees

A detailed treatment of different performance metrics and their calculation methods is presented in section 2.2. However, a common characteristic of PV system performance metrics is that they perform a comparison between the realized energy yield of a plant over a considered period to the expected yield over this period, given knowledge on the solar irradiation incident on the plant. Hence, PV performance metrics lend themselves to be interpreted both as an overall efficiency coefficient as well as a tool to calculate expected yield of the plant at a given irradiance.

As an efficiency coefficient, they are used by plant owners to monitor the performance of their asset over its lifetime. Over time the performance of a plant will decrease due to ageing of the components, called degradation. Performance metrics allow to quantitatively keep track on the rate of degradation over time. Asset owners can use the information on overall degradation of the power plant to inform decisions such as retrofitting or commissioning end-of-life, or start an inquiry in case degradation is occurring at a higher rate than expected. Similarly, they are used by O&M providers to inform smaller-scale maintenance decisions. The maintenance provider can identify underperforming sections of the powerplant by investigating performance metrics, or by comparing solar irradiance measurements directly with other datastreams, at sub-system level.

Performance metrics are also of importance in calculating the financial value of a power plant in case it is sold or the company that owns it is involved in a merger with or acquisition by another company and requires a valuation of its assets. Performance metrics can then be used in combination with an solar irradiance forecast to more accurately calculate the remaining energy yield over the remaining plant lifetime, potentially even including the empirical degradation rate determined using said metrics. The adjusted expected yield is then used in the discounted cash flow calculations on which actual plant valuations are based.

Finally, PV performance metrics are used in the form of contractually specified guarantees in EPC and O&M contracts. As was discussed in section 2.1.2, such contracts specify a guaranteed value of a certain performance metric that must be surpassed

during a post-construction acceptance tests or in annual maintenance service quality assessments. In case the contractually agreed-upon guaranteed value is not achieved, liquidated damages are owed to the asset owner by the EPC or O&M contractor, with an amount typically proportional to the degree of underperformance. This is the application in which context we will study measurement uncertainty, as stated in the introduction. However, a treatment of how uncertainty affects performance guarantees and their conformity assessment is postponed to section 2.4.

Irradiance Forecasts: Yield Assessments, Production Scheduling and Profit Maximization

Irradiance forecasting over long- and short-term are the second important purpose of irradiance measurements.

During the pre-construction phase, irradiance measurements play an important role in site prospecting and the pre-construction energy yield assessment. Prospecting usually comprises of investigating the available solar resource at prospective sites through widely available historical meteorological satellite data. Examples of such (publicly available) datasets and a relative accuracy comparison can be found in AlFaraj et al. (2024). These datasets often contain 15-minute to hourly data of a grid with spatial resolution of a few kilometers stretching back 10-40 years. Once a site is selected, optionally an energy yield assessment is conducted to obtain projections on energy yield thereby cashflow over the project lifetime. While energy yield assessments can be done using exclusively satellite data, including ground data from a (usually 1-year long) field campaign increases the accuracy of the estimate. For further reading on this methodology we refer to Habte et al. (2024), Cebecauer & Suri (2016) and Polo et al. (2016).

The philosophy of pre-construction energy yield assessments is that they allow projected cash flow models to be based on realistic historical data. Providing a factual basis for these models helps the project developer build a bankable business case to be presented to potential investors. However, numerous other factors influence whether a business model is considered bankable: developers are required to provide the financers with detailed technical plant layouts, comprehensive lists of required components, a detailed financial model with motivated inputs (among which is the energy yield assessment), risk mitigation plans, and a full documentation of finalized contracts and permits. Bankability of a business model is not exclusively determined through the energy yield assessment. However, the energy yield assessment is of great significance to the projected cash flow and hence is an important factor that affects *the amount of funding* that can be raised, where higher accuracy forecasts provide lower risk for the financer.

Post-construction, short-term irradiance forecasts are of interest to PV asset managers for day-ahead production scheduling and for intra-day profit maximization. Utility companies typically require an energy production schedule for grid-connected power producing facilities in order to ensure safe and stable grid operations. Deviations from this schedule may cause fines to be imposed on the asset owner. Short-term irradiance forecasts help the plant operator to accurately estimate production in the short term.

The exponential growth of the solar PV industry has resulted in smaller profit margins. PV power is a highly variable source of energy, and PV plants in the same geographical region will experience peak production hours and nighttime hours at (roughly) the same time. This variability in electricity supply has caused a commensurate increase in electricity price volatility, with the prices on the energy market being minimal at times where PV power production is maximal. The drive for profit maximization has led to recent integration of co-located battery energy storage systems (BESS) at PV power plants, that allow to store energy when market prices are low and sell at a later time when prices are high. This has given rise to a novel technological field in which irradiance forecasts are combined with mathematical energy market trading models with the ultimate aim to maximize profit. For further treatment of this topic, see for instance the trading model proposed by Visser et al. (2024) or the study by Campos et al. (2022).

Short-term irradiance forecasting is a developing technological field that employs satellite- and ground-based irradiance data, numerical weather prediction models and more recently machine learning algorithms to obtain accurate forecasts. The recent publication by Di Leo et al. (2025) provides an overview of recent technological advancements in this field.

2.2 Performance Assessment in Utility-Scale Solar PV Power Plants

In section 2.1 we discussed the involved process of development, construction and operations of utility-scale solar PV plants as well as the significant financial interests of various stakeholders. We saw that performance assessment of PV plants plays an important role in post-construction acceptance testing and operations monitoring.

Accurate performance assessment of solar PV power plants is a complicated endeavor which requires defining a suitable performance metric, the acquisition of power, solar and environmental data and a critical analysis of their uncertainty. Standardization agencies have developed performance assessment standards to facilitate accurate and homogeneous performance assessment in the industry. Still, multiple standards exist that define different performance metrics. Furthermore, PV plant owners are not required to employ these standards and may instead opt to use their own standards. This has led to slightly fractious performance assessment practices throughout the industry.

In this section, we will discuss factors influencing PV plant performance, standardized metrics quantifying performance, important quantities to measure for performance assessment, equipment with which to measure these quantities and the uncertainties associated with these measurements. We will devote attention to the most relevant standards for PV system performance assessment and the performance metrics outlined therein.

2.2.1 PV Performance Metrics

General Philosophy of PV System Performance Definitions

Although the approaches taken by different performance assessment methods vary, something they have in common is the philosophy in the question they are trying to answer. All methods are trying to inform the user on *how much energy the powerplant is generating in the current environmental conditions, compared to what can be theoretically expected or achieved in these conditions*.

This means that, in general, performance is defined as measured power output and measured irradiance (power input) compared to some pre-defined benchmark on expected power output at the given irradiance, with possible subsequent corrections for environmental factors. Since most performance metrics include corrections to environmental factors influencing energy output, it is instructive to devote our attention to factors that affect PV system power output before treating the various metrics in detail. Common factors include:

- Module temperature: solar cells become less efficient at converting light to electrical energy as their temperature increases.
- Soiling: dirt or dust on the module surface blocks incoming irradiance.
- Humidity: microscopic water droplets absorb, reflect and diffract light, reducing overall light intensity. This includes fog, dew and frost.
- Spectral response: the efficiency of converting light to electrical energy is dependent on the wavelength of the light and the semiconductor in the PV cell; electrical energy output is dependent on the spectral response of the module as well as the spectral composition of the incoming irradiance.
- Directional response: PV modules are less efficient at converting irradiance from low angles of incidence due to the light being reflected.
- System degradation: degradation is an umbrella term for energy conversion losses due to physical damage to the modules, wiring, inverters or other components over time.

This list is non-exhaustive but contains some of the most important factors influencing power output. A comprehensive review can be found in the publication about this subject by Bamisile et al. (2024).

Standardized PV Performance Metrics

The most widely adopted metrics are defined in the IEC 61724 set of standards and the ASTM E2848 standard, which have distinct approaches. The metrics outlined in the IEC standard are more straightforward and hence will be treated first.

The IEC 61724 standard is comprised of three documents covering a wide range of topics related to PV performance assessment. The 61724-1 document discusses the main performance metric and considerations with regards to performance monitoring. It defines important technical terms, required measurements and measurement equipment, data acquisition, processing and reporting requirements, calculated parameters and performance metrics. The -2 document describes a short-term test to determine the system capacity, including testing procedures and documentation requirements. The -3 documents is similar, but covers a long-term energy yield test. The -1 document provides basic definitions required for the tests outlined in the -2 and -3 documents.

The 61724-1 document contains the definition of the *Performance Ratio*, or PR, which is among the most widely adopted performance metrics. In its basic form, the definition of the PR is given as

$$PR = \frac{\sum_k P_{out,k}\,\tau_k}{\sum_k G_k \tau_k}\frac{G_{STC}}{P_0}. \tag{2.1}$$

In this equation, $P_{\mathrm{out},k}$ is the plant power output, G_k is the incident plane-of-array solar irradiance and τ_k is the duration of measurement interval at time k. The constants P_0 and G_{STC} are the rated plant power output and incident solar irradiance at standard testing conditions (STC). The *standard testing conditions* are a common definition in industry, referring to a condition with an ambient temperature of 25 °C and incident irradiance (i.e. G_{STC}) of 1000 $W \cdot m^{-2}$. The rated DC power of a module is the power it delivers at STC and is specified by the manufacturer. P_0 is the sum of the rated power output of all modules in the power plant.

The PR is in essence a ratio between power output and available solar power at the site, multiplied by a rescaling factor accounting for the efficiency of the modules at standard testing conditions. The advantages of the PR are its straightforward nature and its exclusive dependence on measured data. The IEC 61724-1 norm provides variations on the basic PR that account for temperature considerations or that can be applied to bifacial systems. For instance, the 25 °C temperature-corrected performance ratio makes an adjustment for temperature conditions different from the 25 °C STC temperature. It is given by

$$PR_{25} = \frac{\sum_k P_{out,k}\,\tau_k}{\sum_k C_{25,k} G_k \tau_k}\frac{G_{STC}}{P_0}, \tag{2.2}$$

where

$$C_{25,k} = \gamma\big(T_{mod,k} - 25\big) + 1. \tag{2.3}$$

Here $T_{\mathrm{mod},k}$ is the module temperature at time k and γ is the relative maximum-power temperature coefficient. The variable γ is typically negative and has units of °C^{-1}, so the quantity $C_{25,k}$ is a dimensionless factor that essentially rescales to account for the decrease in module efficiency as its temperature increases.

A survey conducted by the Solar Bankability Project (2016) found that asset managers sometimes apply further corrections to the performance ratio to account for plant availability and degradation during operational lifetime. Assuming a guaranteed plant availability T_{avail} (for instance 99% of the time) and a plant degradation D at the time of calculation, the PR assumes the form:

$$PR_{25,\mathrm{operational}} = \frac{PR_{25}}{T_{\mathrm{avail}}(1-D)}. \tag{2.4}$$

The IEC 61724 standard does not provide a codification for this appended performance metric, nor does it give any indication on measurements to determine the plant degradation.

Similar corrections to the performance ratio can be performed for other loss factors, depending on whether the involved parties want to exclude specific factors from the performance assessment.

Although the PR has many advantages, a critique is that it is not a very subtle metric and that it leaves many influencing factors out of the equation. For instance, instead of simply using the power ratings of the employed PV modules, asset managers often possess elaborate digital models of the power plant that they can employ for performance assessment. The IEC 61724 standard allows to use these models for performance assessment through the second main performance metric: the *Energy Performance Index* or EPI. The definition for EPI is given in the -3 document as

$$\text{EPI} = \text{Measured Energy} \,/\, \text{Expected Energy} = \left(\sum_k P_{\text{out},k} \tau_k \right) / \,\text{Expected Energy} \tag{2.5}$$

The expected energy is defined as the energy yield predicted by the digital model of the power plant, given the measured environmental conditions such as irradiance, temperature, humidity, etc. as inputs. The EPI has as advantage over the PR that it leaves full freedom to include arbitrarily complex considerations to determine the expected energy output. On the other hand, unrealistic models will compare the realized energy output to unrealistic expectations. The EPI is closely related to the last performance metric in the IEC standard: the *Power Performance Index* or PPI, which is the same as the EPI but for the instantaneous measured power and environmental variables instead of over a larger time interval.

A second widely-adopted standard is the ASTM E2848 standard for PV system testing. There is considerable overlap between this and the IEC norm, in the sense that the ASTM norm also prescribes requirements on measurement instrumentation and testing procedures. However, where the IEC standard calculates performance by determining a ratio between realized energy output and an expected energy output given the environmental conditions, the ASTM E2848 standard takes an altogether different approach. The standard prescribes an extrapolation of the power output to reference conditions through fitting the measured environmental variables to it with a multiple linear regression model, which is subsequently used to calculate a fitted reference power output using reference environmental conditions as input. Mathematically:

$$P = G(a_1 + a_2 G + a_3 T_a + a_4 \nu), \tag{2.6}$$

$$P_{RC} = G_{RC}(a_1 + a_2 G_{RC} + a_3 T_{RC} + a_4 \nu_{RC}). \tag{2.7}$$

The coefficients a are the regression coefficients, G and T_a are the POA irradiance and ambient temperature, respectively, and ν is the wind speed. The subscript RC stands for 'reference conditions'. The extrapolated power output can then be compared to an expected power output at reference conditions that is calculated by a computational model of the power plant.

The advantage of this method is in the fact that it incorporates environmental factors like temperature and wind speed at baseline and is not dependent on fixed reference conditions, since these are allowed to be specified by the user. However, it is reliant upon an assumption that power output at specified reference conditions can be linearly extrapolated from measured conditions, the validity of which is debatable.

2.2.2 Measurements for Performance Assessment

Performance calculations of PV assets require gathering a large amount of data on the environmental and operating conditions of the plant. The standards just discussed establish strict requirements on quantities to measure, accuracy bounds, measurement equipment and even on calibration and maintenance of said equipment. Often, the standards will refer to

other standards in specifying these requirements. This section is a brief overview of relevant quantities that are measured, ways in which they are measured and standards relating to these measurements. An emphasis is placed on the uncertainty of the measurement equipment.

Before we continue, it is important to understand that the standards discussed in this section often define an ordinal classification system with increasingly stringent requirements. For instance, the IEC 61724-1 document outlines class A and class B monitoring systems, where class A systems are more accurate and comprehensive but require a greater investment on the user's part. In this section we will treat all classes outlined by a standard where applicable.

Power Output

Continuous output measurements are required to estimate the yield of the PV system under consideration. PV modules capture solar radiation which causes a voltage, that in turn drives a DC current. This DC current passes through a transformer to tune the IV characteristic for optimal power production. Afterwards, the DC current is fed into an inverter that transforms it to AC current. The combined AC current of the entire power plant is then delivered to the grid.

Plant design varies on where the current generated by the strings is aggregated before delivery to the grid. In *centralized inverter* design, all generated DC currents are combined and fed into a single large inverter that delivers a single AC output. While economical due to the expensive nature of inverters, this design has an obvious vulnerability with respect to the availability of this singular inverter. In *string inverter* design, the DC output current is converted to AC by individual inverters at string-level, and delivered to the a plant-level AC grid, from which it is subsequently delivered to the central grid. This has the advantage that there is no crucial dependence on a single component, but comes at a greater monetary investment due to the expensive nature of inverters. For further reading on the subject of powerplant design we refer to Villacorta and Victoria (2024).

The term 'power generated by a solar PV system' merits some deliberation: is it the DC power fed into the inverter, is it the AC power that is delivered by the inverter or is it the AC power delivered to the grid? Power losses occur during the journey from the modules to the grid: resistive cabling losses and inverter losses most prominently. The location at which power output is defined dictates which are part of the system, and which should be accounted for.

The ASTM E2848 norm leaves this choice to the user, while the IEC 61724 norm explicitly defines the use of AC power in performance assessment. In this thesis we will understand output power as the total AC output power of a plant, as in the IEC standard.

The IEC 61724 standard defines the power metering equipment to be of class 0,2 S for class A systems and class 0,5 S for class B monitoring systems, which are metering accuracy classes defined in IEC 62053-22 (2020).The maximum permissible measurement uncertainty according to this standard depends on the instantaneous operating conditions of the system. The visualization of the error bounds is presented in figure 2.3 for a power factor of 1 and an operating range between 0 and 10 Ampère. The standard defines stricter uncertainty limits once the current exceeds 5% of the maximal current, with the 0,2S class allowing a power metering uncertainty of $\pm 0.2\%$ in this case. Slightly different specifications are given for non-unity power factors, but power plant operators ordinarily aim for a unity power factor since this is most efficient.

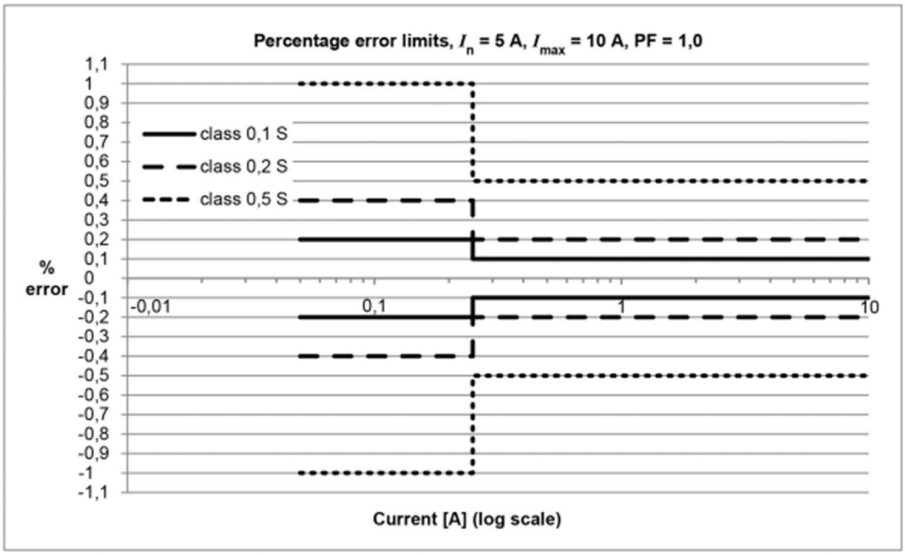


Figure 2.3: Uncertainty specifications of different metering accuracy classes in IEC 62053-22 (2020). Accuracy specifications differ depending on whether the current exceeds 5% of the system's maximally allowed current (seen in figure) and depending on the power factor. Error bounds shown are for unity power factor, which corresponds to optimal operating conditions.

The ASTM E2848 norm defines the maximum uncertainty in output power to be at most 1.5% of the output power at reference conditions. This is significantly less strict than the requirements of IEC 61724.

Solar Irradiance

Solar irradiance is the resource that is converted to energy by the power plant. Knowing how much of this resource is available for energy generation is crucial for performance assessment. However, measuring this resource is considerably less straightforward than measuring power output. Therefore, some crucial physical distinctions must be made.

First, we distinguish between irradiance and irradiation. Irradiance is the amount of electromagnetic radiation irradiation a specific surface area per second, measured in $W \cdot m^{-2}$. Irradiation is the total amount of irradiance received by a surface area over a certain timeframe, and is hence measured in $J \cdot m^{-2}$, $kWh \cdot m^{-2}$ or equivalent.

Secondly, there are three types of irradiance that can be referred to: global, direct and diffuse. Global irradiance is all irradiance incident on a surface, which is the sum between direct and diffuse irradiance. Direct irradiance is the irradiance incident on a surface that has traveled from the light source (in our case the sun) to the surface without being scattered along the way. Conversely, diffuse irradiance is all irradiance which is not incident directly from the source but is instead scattered onto the surface by surrounding objects or molecules in the atmosphere.

As a crude rule-of-thumb, on a clear-sky day diffuse irradiance comprises about 15% (Spitters et al., 1986) of global irradiance (in the 300-3000 nm wavelength range). In overcast conditions this fraction is obviously higher and approaches 100%. Additionally, while direct irradiance always has the same spectral decomposition, the spectral decomposition of diffuse irradiance is dependent on the (color of the) direct environment, cloud conditions, humidity, aerosol depth and solar zenith angle.

Each of these three irradiances (global, direct and diffuse) require different instruments to measure. Direct irradiance is measured using a pyrheliometer, which is a detection device with an extremely narrow field of view (standardized at 5° by the world meteorological organization). The pyrheliometer is mounted on a tracker system which ensures it is pointed directly at the sun.

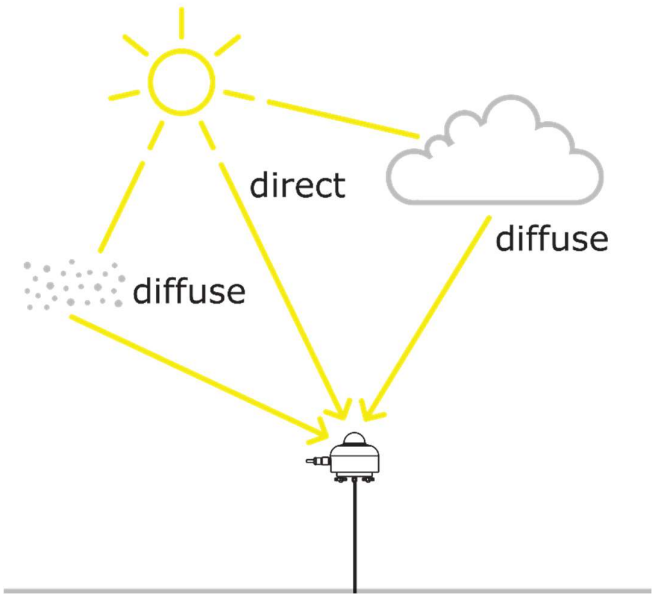


Figure 2.4: Difference between direct and diffuse irradiance as shown by Hukx. Global irradiance is the sum of direct and diffuse irradiance. Diffuse irradiance also encompasses irradiance reflected by the ground or nearby physical structures.

For the purpose of PV performance testing, global irradiance can be measured in two different ways: either by using a pyranometer or using a PV reference cell. The pyranometer is a specialized instrument for measuring all irradiance in a hemisphere (i.e. it has 180° field of view), while a PV reference cell is simply a small PV module for which the power output is characterized against incident irradiance. PV reference cells are less accurate than pyranometers and moreover not spectrally flat, but are much cheaper than pyranometers.

Diffuse irradiance can also be measured in two different ways. Historically, diffuse irradiance was measured by employing a regular pyranometer with a fixture attached to it that blocks out direct solar irradiation, like a shading ball or a rotating shadow band. These fixtures have to be manually adjusted at regular intervals to account for seasonal changes in solar altitude. Modern diffusometers are dedicated instruments that filter out direct irradiance in a passive manner, for instance through an internal detector array that is shaded by a built-in shading mask.

Although knowledge of the direct/diffuse split can also be of influence due to the different spectral composition of diffuse light, global irradiance is ordinarily of greatest interest for PV system performance analysis. Here, we will make another important distinction between *global horizontal irradiance* (GHI) and the *plane-of-array (POA) irradiance*, which is the global irradiance in the plane in which the PV modules are oriented.

Pyranometers typically consist of a weather-resistant body with a hemispherical glass dome on top, underneath which is a black-coated thermopile. Incident radiation passes through the dome (which filters the light to the desired bandwidth) and is absorbed by the thermopile, which will heat up. The thermopile coating is chosen such that it has a spectrally flat response in the range 300-3000 nm. The heating of the thermopile causes a temperature difference to arise between it and the pyranometer housing, which is converted to a small electrical signal through the thermoelectric effect. This electrical signal is proportional to the heat difference between the thermopile and the pyranometer housing, which is in turn proportional to the amount of incident irradiance on the thermopile.

The pyranometer measurement equation thus reads as follows:

$$G = \frac{\mathrm{V}}{\mathrm{S}} \tag{2.8}$$

In this equation, G is the total irradiance in a hemisphere along the pyranometer tilt axis in units of $W \cdot m^{-2}$, V is the voltage output in Volts and S is the pyranometer sensitivity, in $\frac{V}{W \cdot m^{-2}}$.

Newly built pyranometers have to be calibrated against a reference instrument before they can be used in order to know the sensitivity constant S. Usually this calibration is done against a reference pyranometer, although other methods using pyrheliometers exist as well. Calibration procedures have been codified in the ISO 9847 standard (ISO, 2023). For class A pyranometers, this standard lists achievable calibration uncertainties of 1%. Pyranometer sensitivity will drift slowly over time, requiring it to be recalibrated at regular intervals, typically every two years.

Pyranometer measurements are subject to several sources of uncertainty. The ISO 9060 standard lists these uncertainties and specifies tolerance intervals for each uncertainty source (ISO, 2018). Pyranometers are classified according to their compliance to class uncertainty limits. Common pyranometer error sources and their accepted magnitudes by class are presented in table 2-1, which is taken from the ISO 9060 standard.

Table 2-1: Pyranometer class and associated accepted error specifications as defined in in the ISO 9060 pyranometer standard (ISO, 2018).

Parameter	**Name of classes, acceptance intervals and width of the guard bands (in brackets)**		
Name of the class	**A**	**B**	**C**
Response time: time for 95% response	< 10 s (1 s)	<20 s (1 s)	<30 s (1 s)
Zero off-set: a) Response to -200 W·m^{-2} net thermal radiation b) Response to 5 K·h^{-1} change in ambient temperature c) Total zero off-set including the effects a), b) and other sources	 ±7 W·m^{-2} (2 W·m^{-2}) ±2 W·m^{-2} (0.5 W·m^{-2}) ±10 W·m^{-2} (2 W·m^{-2})	 ±15 W·m^{-2} (3 W·m^{-2}) ±4 W·m^{-2} (1 W·m^{-2}) ±21 W·m^{-2} (2 W·m^{-2})	 ±30 W·m^{-2} (3 W·m^{-2}) ±8 W·m^{-2} (1 W·m^{-2}) ±41 W·m^{-2} (3 W·m^{-2})
Non-stability: Percentage change in responsivity per year	±0.8 % (0.2 %)	±1.5 % (0.2 %)	±3 % (0.5%)
Nonlinearity: Percentage deviation from the responsivity at 500 W·m^{-2} due to change in irradiance within 100 W·m^{-2} to 1000 W·m^{-2}	±0.5 % (0.2 %)	±1 % (0.2 %)	±3 % (0.5 %)
Directional response (for beam radiation): The range of errors caused by assuming that the normal incidence responsivity is valid for all directions when measuring from any direction (with an incidence angle for up to 90° or even from below the sensor) a beam radiation whose normal incidence irradiance is 1000 W·m^{-2}	±10 W·m^{-2} (4 W·m^{-2})	±20 W·m^{-2} (5 W·m^{-2})	±30 W·m^{-2} (7 W·m^{-2})
Clear sky global horizontal irradiance spectral error: Maximum spectral error observed for a set of global horizontal irradiance clear sky spectra, see ISO 9060 A.7 for calculations.	±0.5 % (0.1 %)	±1 % (0.5 %)	±5 % (1 %)

Temperature response: Percentage deviation due to change in ambient temperature within the interval from -10 °C to 40 °C relative to the signal at 20 °C	±1 % (0.2 %)	±2 % (0.2 %)	±4 % (0.5 %)
Tilt response: Percentage deviation from the responsivity at 0° tilt (horizontal) due to change in tilt from 0° to 180° at 1000 $W \cdot m^{-2}$ irradiance	±0.5 % (0.2 %)	±2 % (0.5 %)	±5 % (0.5 %)
Additional signal processing errors	±2 $W \cdot m^{-2}$ (2 $W \cdot m^{-2}$)	±5 $W \cdot m^{-2}$ (2 $W \cdot m^{-2}$)	±10 $W \cdot m^{-2}$ (2 $W \cdot m^{-2}$)

The uncertainty bounds are mostly straightforward, except for the directional response error. The uncertainty bound for this error is defined as the maximum permissible error for a beam coming from any direction with a normally incident component of 1000 $W \cdot m^{-2}$. In other words, the maximum permissible directional response is dependent on the amount of incident irradiance.

More specifically, in case the direct component of the global irradiance is E, and the solar zenith angle is θ, the dynamical error bound a for the directional response error is the specification limit b multiplied by a factor:

$$a = b \cdot \frac{E}{1000\ Wm^{-2} \cdot cos(\theta)}. \tag{2.9}$$

According to Konings and Habte (2015), E can to a good approximation be replaced by the measured global horizontal irradiance in case no direct solar irradiance measurements are available.

Full pyranometer uncertainty assessments following GUM methodology have been performed by Konings and Habte (2015) and Habte et al. (2024), and a similar calculation is provided with the ASTM G213-17 (2017) standard. Uncertainties between 2-5% for daily totals are common for class A pyranometers. Figure 2.5 shows an uncertainty analysis for pyranometer measurements during a clear-sky day, with the uncertainty split by source.

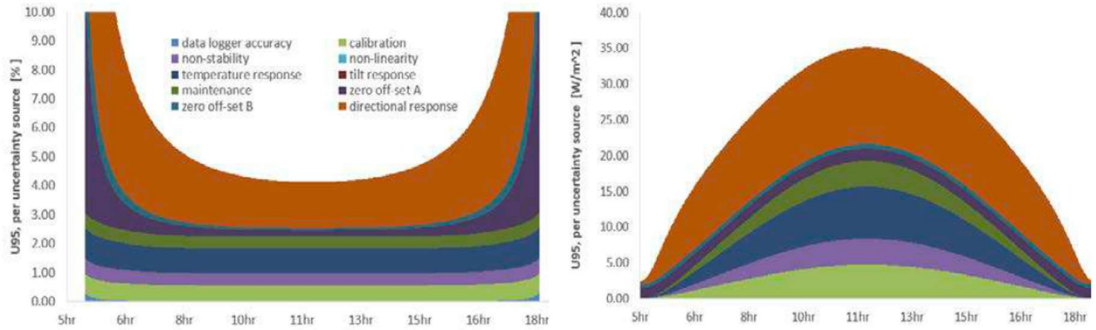


Figure 2.5: Pyranometer uncertainty split by source for a single day. The left figure shows uncertainty relative to the measured global horizontal irradiance. The right figure shows the uncertainty in absolute terms. Figure taken from Habte et al. (2024).

As mentioned, another method to measure global irradiance is by employing PV reference cells, essentially miniature PV modules for which the power output is calibrated at STC. PV reference cells are much cheaper than thermopile pyranometers, but this comes at a tradeoff in device capabilities. First of all, PV reference cells do not actually measure global irradiance since their measurement sensitivity is not spectrally flat. Rather, they have the spectral responsivity of the used PV module type. Further drawbacks of PV reference modules are the fact that their responsivity is dependent on module temperature, their high sensitivity to reflection-induced errors at low incidence angles and their sensitivity to soiling as a consequence of

their flat surface. However, PV reference cells have lower response times compared to pyranometers. The IEC 60904-2 (2023) document provides a standard for PV reference cells and includes physical device requirements and characterization procedures.

Due to these considerations in the previous paragraph, pyranometers are better suited for measuring global irradiance than PV reference cells. However, it has been argued that PV reference cells are better suited to estimating PV system performance since they exhibit the exact same characteristics as the system, if fitted with a PV module of the same type as the system modules. That is: they have identical temperature, directional and spectral response as the PV modules. Indeed, a study performed by Meydbray et al. (2012) found that PV reference cells have lower uncertainty in measurements of "PV irradiance" than pyranometers, in which "PV irradiance" is a specific irradiance quantity corresponding to global irradiance weighted by the spectral and directional responsivity of the PV module under consideration. In this study, pyranometer uncertainty was found to be two times greater than PV reference module uncertainty in measuring PV irradiance.

It is important to understand the implications of this redefinition of global irradiance and the implications its usage has in PV performance assessment. By defining performance as energy output with respect to the available "PV irradiance" input, one assesses how well the PV system is converting the effective available energy to electricity. This can help identify performance losses due to factors like soiling and module degradation in a more direct way (provided that the PV reference module itself is not soiling or degrading at the same rate as the system).

However, the above method invariably ensures higher performance, since other loss factors like directional response are ignored. This, crucially, is also why one must be careful with redefining performance in this way. By defining the available energy as the effective energy that the PV system can convert, we apply an artificial rescaling that reduces the importance, or even outright ignores, a subset of the available energy. Thus, we are no longer assessing how well the system is capable of converting available solar energy to electrical energy. It neglects, for instance, the unrealized energy output due to capture losses of light incident from low angles or from wavelengths that the used module is not sensitive to. Likewise, usage of low-quality modules with low responsivity over the entire spectrum will not be detected in performance calculations, since the measured solar resource is rescaled the same way. The interpretation therefore gives a false indication of plant performance and may negatively influence decisions by masking loss sources.

The IEC 61724-1 document specifies an increasing number of pyranometers to be placed on-site depending on the size of the plant. They are furthermore required to be placed at representative locations throughout the power plant. The standard requires for both global horizontal and plane-of-array to each be measured by at least 2 pyranometers for plants with capacity less than 40 MW, 3 pyranometers for plants between 40-100 MW and an additional pyranometer for each additional 200 MW added capacity.

The ASTM E2848 norm does not specify a minimal number of pyranometers that must be used or what their uncertainty limits are allowed to be. The standard refers to ASTM G183-15 (2015) for pyranometer characteristics influencing uncertainty, which in turn recognizes the ISO 9060 norm for uncertainty classes. ASTM 2848 requires a pyranometer (of any class) to be used by default, but does allow for use of a PV reference device in case all parties to the test agree on this.

Temperature

There are two temperature measurements that are relevant for PV system performance assessment: the ambient air temperature and the PV module temperature. In section 2.1.3 it was discussed that the efficiency of PV modules decreases as their temperature increases, and we have seen that PV performance assessment methods provide ways to correct for this by explicitly including temperature in the performance metric. The temperature-corrected performance ratio PR_{25} includes the module temperature, while the ASTM E2848 regression model includes ambient temperature. Module temperature is usually measured using a dedicated sensor, but the IEC 61724-2 document also provides a formula to estimate module temperature from environmental measurements through the following equation:

$$T_{\text{mod}} \approx G_{POA} \cdot \left(e^{a+b \cdot v_W}\right)[°C\ \ m^2/W] + T_a, \tag{2.10}$$

where v_W is the wind speed and a and b are module-specific constants that are empirically determined. Note that this is a rough approximation with no quantified uncertainty.

ASTM E2848 only prescribes measuring ambient temperature, while IEC 61724 prescribes measuring both ambient and back-of-module temperatures. The standards have identical specifications on the uncertainty limits for the temperature sensors: resolution $\leq 0.1°C$ and total measurement uncertainty less than $\pm 1°C$.

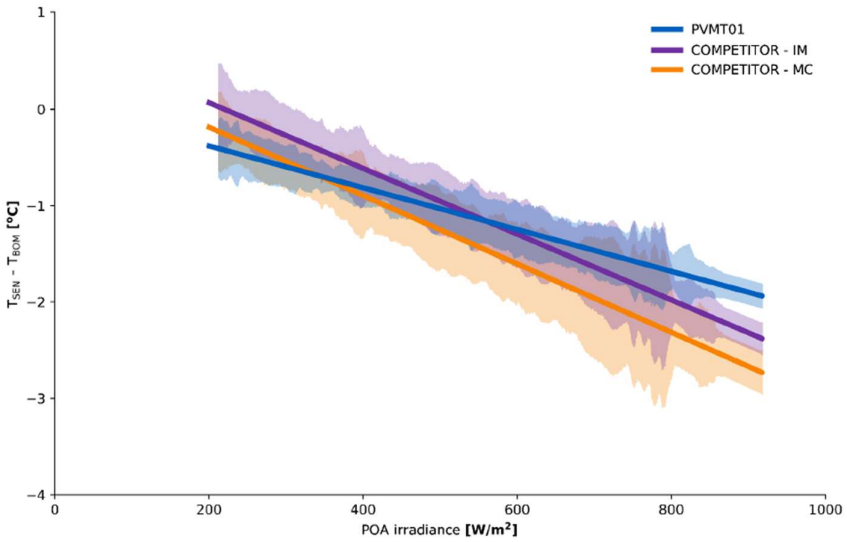


Figure 2.6: Temperature reading and uncertainty compared to actual module temperature for different back-of-module temperature sensors at different irradiance levels. A systematic underestimation as irradiance (and hence module temperature) increases can be clearly discerned from the data. Figure provided under copyright by Hukx.

Although temperature sensors are beholden to strict resolution and measurement uncertainty requirements, module temperature sensors often contain an implicit error in their measurement. An internal study performed by Hukx showed that module temperature sensors tend to underestimate the module temperature by up to $2.5°C$, as can be seen in figure 2.6. The study compared the temperature readings of various dedicated back-of-module temperature sensors to the actual module temperature for different incident irradiance levels. Naturally, as the irradiance level increases the module temperature increases as well, which causes a temperature difference between the module temperature and the ambient air temperature. Hence, the back-of-module temperature sensors will measure a temperature somewhere in-between. Despite this fact, Hukx does not recommend insulating the temperature sensors since this will prevent thermal energy from leaving the sensor, causing an overestimation of the module temperature.

Thus, while standards restrict temperature uncertainty to $\pm 1°C$, the uncertainty may realistically be a few degrees higher than this specification.

Soiling

Soiling refers to the deposition of soil, dust, frost, debris, or other physical material deposited on the surface of PV modules and domes of pyranometers, blocking incident light. Soiling is a common and significant loss factor of PV system yield, requiring regular maintenance of the modules to keep at acceptable levels. If left unchecked in dust- or pollution-heavy environments, it has been shown that soiling can reduce the yield of PV modules upwards of 50% (Warade & Kottantharayil, 2018).

Although rainfall helps clean PV modules and pyranometers, regular cleaning is required to keep the yield loss due to soiling to a minimal, or at least acceptable, level. The frequency with which cleaning is warranted depends on the soiling rate in the local environment. The amount of soiling on a PV module is dependent on a wide variety of factors, including (but not limited to) the local environment, pollution levels, humidity, rainfall, module tilt angle and module surface material. The rate of soiling on a flat surface in a dry, dusty environment with little rain is very different than on a tilted plane in a temperate environment with frequent rain.

The report by Reise et al. (2018) lists estimations on annual PV yield loss ranging between 0.5-4% for systems in different climates with regular cleaning regimes. Additionally, studies have been performed with the aim to characterize the soiling rate of PV modules and pyranometers. A compilation [1] of references to published studies reporting experimentally determined soiling rates of PV modules at various locations worldwide is maintained by the National Centre for Photovoltaic Research and Education of the Indian Institute of Technology Bombay.

While module soiling decreases energy yield and hence negatively affects performance, it does not affect the accuracy of the performance assessment as a whole. Philosophically, we defined performance as the degree to which a PV system utilizes available light for energy generation. Thus we want to compare the true power output to the true available solar irradiance, and the uncertainty on the output side is determined by the accuracy of the power metering equipment. Following this logic, soiling of pyranometers *does* affect the accuracy of the performance assessment, since soiling of the dome induces a deviation between the pyranometer reading and the true irradiance. We must therefore take the effect of pyranometer soiling on irradiance measurements into account.

The argument can be raised that this approach is inconsistent: what merits the exclusion of blocked irradiance due to cloud cover or atmospheric depth as a loss factor while soiling is not excluded? Indeed, both are the result of irradiance being physically blocked from reaching the PV module. Simply put, we recognize cloud cover and atmospheric depth as factors beyond the control of asset owners, while soiling losses can be controlled through maintenance. We distinguish between environmental factors dictating available irradiance, and losses in converting this to electrical power due to the physical state of the PV system. Losses due to self-shading of PV arrays and other controllable factors at design-level are also included in the latter category.

This philosophy is in line with the definition for class A monitoring systems per the IEC 61724 standard. For class B precision, soiling is considered part of the weather. Since we aim to quantify the uncertainty in state-of-the-art performance estimation, it is natural to follow class A definitions.

The degree of soiling is typically quantified through the soiling ratio SR, which is defined as the ratio between the output of the measurement device under consideration and the output of an unsoiled (identical) reference device (Fuke & Kottantharayil, 2025),

$$SR = \frac{\text{Signal}_{\text{soiled}}}{\text{Signal}_{\text{unsoiled}}}. \tag{2.11}$$

The ‘signal’ refers to the irradiance measured by the instrument. Following this definition, Fuke and Kottantharayil (2025) found a soiling rate between 0.19-0.29%/day near Mumbai, India. This study notably does not report rainfall. Other studies, in which rainfall was not controlled, conducted in Qatar reported soiling rates between 0.22-0.53%/day (Bachour et al., 2016). A study by Waters et al. (2018) conducted in an arid region of the US obtained a result of -0.073%/day during an extended period of no rainfall.

[1] Available at https://www.ncpre.iitb.ac.in/ncpre/pages/seriius-soiling-rate-of-the-world.php (accessed 07 January 2026).

Accurate performance assessment requires regular cleaning of pyranometers. The IEC 61724 standard specifies weekly inspection of front-side sensors. The above mentioned literary values aggregate to 1.33-2.03% (Fuke & Kottantharayil,2025), 1.54-3.71% (Bachour et al., 2016) and 0.51% (Waters et al., 2017) signal loss after 7 days.

Due to the high degree of variability in soiling estimates and the high dependency on regional and local conditions, this work will exclude uncertainty due to soiling. It is important to be aware of the significant hit to the measurement accuracy that can be caused by soiling, but due to its complex nature this work will assume measurements performed using clean pyranometers. A methodological assessment of uncertainty due to soiling in different climates and cleaning regimes is an important topic for further study.

2.2.3 Limitations of Performance Metrics

Solar PV has seen a swift evolution from being a niche power generation method to becoming a large-scale industry over the past decades, and with it has come a commensurate maturing of the industry as best practices are codified in a multitude of standards on a wide range of relevant topics. There are some limitations and topics of contention in contemporary standards. We will highlight those related to performance assessment in this section.

First of all, current performance assessment methods have difficulty including periods of clipping and curtailment. Clipping is the reduction of power output due to the power delivered by the modules exceeding the rated inverter capacity, while curtailment is the reduction of power output ordered by the grid operator due to high grid loads. During such periods, the plant is operating at a lower capacity than the modules can achieve by purposeful reduction of power output by the transformers (the maximum power point trackers). Although the two terms are similar, there are subtle differences which merit separate treatment.

During periods of curtailment the grid operator forces the plant to operate at less-than-maximum capacity (given the current solar irradiance), which means the plant is essentially operating at sub-optimal performance. Hence, it would not be an honest representation to include this period in performance assessment, as it artificially decreases the overall performance. However, periods of curtailment often occur at times when the powerplant is delivering a lot of energy to the grid. Hence a period of theoretical high performance is excluded, increasing the relative weight of periods of lower performance and again artificially decreasing the overall performance. Standards specify the exclusion of periods of curtailment from the performance assessment.

In the case of clipping the reduction in power output is caused by insufficient inverter capacity. Instead of an external order to artificially lower production, during clipping the physical output limit of the plant is reached. Hence, a good case can be made that this means a lower performance is merited, as it is attributable to plant design decisions. Due to high inverter costs, PV plants designers make a tradeoff between inverter capacity and the missed income due to energy clipping, choosing for the less costly option. The IEC 61724 standard considers clipping a performance loss and therefore includes these periods by default, while the ASTM E2848 standard explicitly excludes periods of clipping.

A second inherent limitation of PV performance assessment is the continuous effort it takes for the responsible operator to conduct. The measurement equipment required for the acquisition of relevant data requires regular maintenance and recalibration. Pyranometers require recalibration every two years, which means they must be uninstalled and shipped to a calibration lab. This results either in data unavailability, or requires a back-up pyranometer to be placed on-site temporarily. Pyranometers moreover require regular cleaning due to soiling and dust deposition on the dome. The frequency at which cleaning is required is highly dependent on environmental conditions and air quality, but in polluted environments the irradiance measurement error due to soiling can be as high as 3.7% after one week (Bachour et al., 2016).

Another important aspect to highlight is the manipulability of the performance assessment. For instance, by not properly cleaning pyranometers, soiling can be used as a method to artificially inflate performance. A soiled dome will cause a

systematic reduction in registered irradiance and therefore will lead to an artificially increased measured performance. Other such methods to influence performance include the intentional misalignment of a pyranometer to a lower angle than the plane-of-array or by insulating module temperature sensors to hamper heat dissipation and thereby increase its temperature reading.

Involved parties with malicious intent can employ such methods to influence the performance assessment of solar PV systems to their advantage. Contemporary standards require logs to be made of system maintenance, and the degree of detail in such logs may limit the ability to employ such methods as mentioned above. However, no best practices on monitoring the impartiality of the performance assessment process have been published as of yet.

2.3 Uncertainty and Decision-Making

"*Any measurement you take without the knowledge of its uncertainty is completely meaningless.*"
-Walter Lewin

Uncertainty is intrinsic to all measurements. Performance assessment is therefore also an inherently uncertain endeavor. In section 2.2.2 we discussed various quantities involved in performance assessment and uncertainty specifications for their measurement given the IEC and ASTM standards. These standards make no comment, however, on overall uncertainty bounds or on how uncertainty ought to be included in decisions. In this section we will discuss general theory of uncertainty, achievable uncertainties for the discussed performance metrics as presented in academic discourse and guidelines on the inclusion of uncertainty in decision-making processes.

Before we discuss common uncertainties in performance metrics, we must understand how to handle uncertainty: a subtle field of mathematics in which one must take care not to make mistakes. Thankfully, the Guide to the Expression of Uncertainty in Measurement (Joint Committee for Guides in Metrology [JCGM], 2008), referred to as the GUM, provides instructions on how to correctly handle measurement uncertainty and propagate it to derived quantities. We will briefly overview the most important theoretical concepts, but we will leave a detailed treatment of uncertainty propagation to chapter 3.

2.3.1 Basics of Measurement Uncertainty Propagation

To start, we state some important definitions. First of all, we make a distinction between the *error* and the *uncertainty*:

- The *error* is the difference between a measured or estimated value and the unknowable true value of the quantity under consideration.
- The *uncertainty* is a quantification on how close an obtained estimate is to the true value of a quantity.

Moreover, we must make a distinction between different types of error sources:

- *Measurement errors* are the result of the finite accuracy of measurement equipment. Measurement uncertainty due to these errors is usually quantifiable and specifications are often provided by the equipment manufacturer.
- *Model errors* are the result of using models which inaccurately represent reality. These are often much harder to quantify and consider in the total uncertainty evaluation.

Finally, we must make a distinction between two types of uncertainty:

- *Type A uncertainties* are uncertainty quantifications obtained by estimation from repeated measurements.
- *Type B uncertainties* are obtained by evaluating instrument or calibration specifications, through physical modelling or obtained from sources. These types of uncertainty have a specified distribution.

The GUM provides a systematic framework for propagating uncertainty to derived quantities. In GUM terminology, a derived quantity $Y = f(X_1, \dots, X_N)$ that is a function of input quantities X_i is called a *measurement model.* We will discuss the mathematics in detail in chapter 3, but the general framework to calculate the uncertainty of a measurement model is quite straightforward. First we obtain uncertainty estimates for all dependencies X_i, which requires a different treatment for type A and type B sources (outlined in the GUM). Next, the sensitivity to each uncertainty is obtained by taking the partial derivative of f with respect to its uncertain dependencies. The uncertainties are rescaled by their respective sensitivities and subsequently combined through a root-sum-square method that allows to account for correlation between uncertainties to obtain the total uncertainty in Y.

The GUM methodology is generally not an exact framework for calculating uncertainty of derived quantities since it relies upon Taylor expansion to estimate the sensitivity coefficients. Moreover, while correlation between uncertainty sources can be included in the framework, it is not a trivial task to obtain a quantification on the exact nature of the correlation between two uncertainty sources which makes it hard to include in practice. Therefore, uncertainties in derived quantities calculated using the GUM are estimates themselves. In practice these estimates are deemed sufficiently accurate, and no uncertainty quantification is given for the error in the uncertainty quantification.

2.3.2 Representing Uncertainty

In order to make decisions based on uncertain information we must employ clear terminology to avoid making mistakes. Generally, uncertain quantities are reported as $\mu \pm 2\sigma$, with μ the measured value and the uncertainty σ. If no other information is accompanying this reporting, one may assume that this defines a normal distribution with mean μ and standard deviation σ for the true value of the quantity under consideration.

An important topic when using uncertainty is the *confidence interval*, which are intervals in the space of real numbers that are likely to contain the true value of the quantity with a certain *confidence level*. For instance, the interval $(-\infty, +\infty)$ is a confidence interval of 100%, since this interval covers all possible numbers. We can construct confidence intervals for the normal distribution by using its two parameters: the interval $[\mu - \sigma, \mu + \sigma]$ is a 68.27% confidence interval, while $[\mu - 2\sigma, \mu + 2\sigma]$ is a 95.45% confidence interval. Figure 2.7 shows confidence intervals between different multitudes of the standard deviation for the normal distribution.

A measurement and its uncertainty may be presented with an accompanying *coverage factor* k. If a quantity is presented as $\mu \pm \zeta\ (k = \kappa)$, this means that $\zeta = \kappa\sigma$. Coverage factors are often used to provide readers with a confidence interval for the result of a higher confidence level than the 68.27% of the interval $[\mu - \sigma, \mu + \sigma]$ that is readily constructed if σ is reported as uncertainty. By instead reporting the *expanded uncertainty* ζ and stating a factor of 2, the user can immediately construct a 95.45% confidence interval around the measured value. For a strict 95% confidence interval one must use a coverage factor of 1.96.

Coverage factors for arbitrary confidence levels can be obtained by reading out the quantile function of the normal distribution, which is called the *probit* function. For an 80% confidence interval, one can obtain the required k-value by evaluating $Probit(0.1)$ or $Probit(0.9)$. More generally, the *quantiles* of a probability distribution are a set of points dividing the total range of possible outcomes in segments (confidence intervals) of equal probability. Some special quantiles have their own names: 25%-quantiles are called quartiles, 10%-quantiles are called deciles and 1%-quantiles are called percentiles.

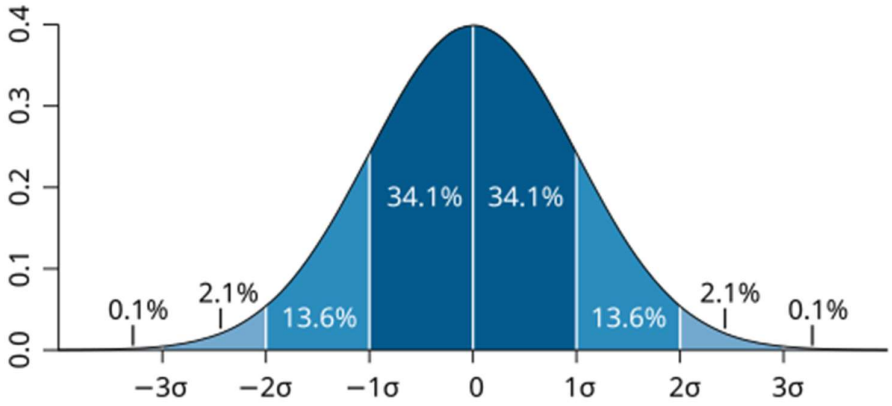


Figure 2.7: Figure illustrating the confidence intervals of a realization of a normally distributed random variable to fall between multitudes of the distribution standard deviation. Image source: Wikipedia[2].

A final definition from probability theory which is of great importance to this study is the concept of *exceedance probability levels,* which is closely related to quantiles and confidence intervals. The most well-known of these is the *median* of a probability distribution, which is the value that has a 50% probability of being exceeded. The median is therefore often referred to as the P50. Similarly, the P90 is the value that has a 90% probability of being exceeded. Observe that the P50 and P90 correspond to the 50th and 10th percentile, respectively.

The P50, P90 and similar are of great significance in decision-making because they characterize one-sided confidence intervals at a specified confidence level instead of the two-sided $\mu \pm k\sigma$ intervals. Their relevance is best illustrated through an investment decision: suppose we know the probability distribution of the expected return on an investment. To be safe, we want to ensure that there is a 90% probability we do not lose money on this investment. Thus, the investment may not cost more than the P90 value of the projected returns. P50 and P90 values of projected cash flows are often included in business proposals as 'best estimate' and 'high certainty' projections that allow investors to make informed decisions, rather than reporting a best estimate with symmetric uncertainty bounds.

2.3.3 Reducing Uncertainty through Multiple Measurements

An instinctual way to attempt to increase the certainty in a measured value is by repeating the measurement and taking the average of the two measurements. Indeed, performing repeated measurements of a stochastic variable and averaging the outcomes results in the average converging to the expected value of the variable's distribution, with the uncertainty tending towards 0, as a consequence of the *law of large numbers*. It is instructive to illustrate this point with a concrete example. Combined uncertainties provided in this section are calculated following the GUM framework for independent uncertain variables, a detailed treatment of which is deferred to chapter 3 and appendix A.

Assuming the mathematical ideal that two measurements of the same variable can be performed independently from each other, we can obtain two measurements μ_1, μ_2 of this variable with uncertainties σ_1, σ_2. We can then average this into a better estimate of the variable:

[2] Standard deviation diagram showing a normal distribution. Wikipedia contributors, *Wikipedia* (https://en.wikipedia.org/wiki/Normal_distribution). CC BY-SA 3.0.

$$\frac{\mu_1 + \mu_2}{2} \text{ with uncertainty } \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}. \tag{2.12}$$

Assuming the uncertainty over repeated measurements is identical, after N repeats we have reduced the uncertainty by a factor $\frac{1}{\sqrt{N}}$. While adjusting the estimate of the measured variable by taking the average is intuitive, it is not necessarily mathematically optimal in case some measurements are made with greater accuracy than others. Instead, a greater weight should be attributed to more certain measurements through an expression of the form

$$\frac{w_1\mu_1 + w_2\mu_2}{w_1 + w_2} \text{ with uncertainty } \sqrt{\frac{w_1^2\sigma_1^2 + w_2^2\sigma_2^2}{w_1^2 + w_2^2}}. \tag{2.13}$$

The weights that provide an optimal (unbiased) estimator of the measured variable are given by *inverse variance weighting*, i.e. setting $w_i = \frac{1}{\sigma_i^2}$. In figure 2.8, the difference in the uncertainty of the measurement fusion through regular averaging and inverse variance weighting can be seen clearly. The above expression is generalized for multivariate measurement models in the form of Kalman filtering (Gibbs, 2011).

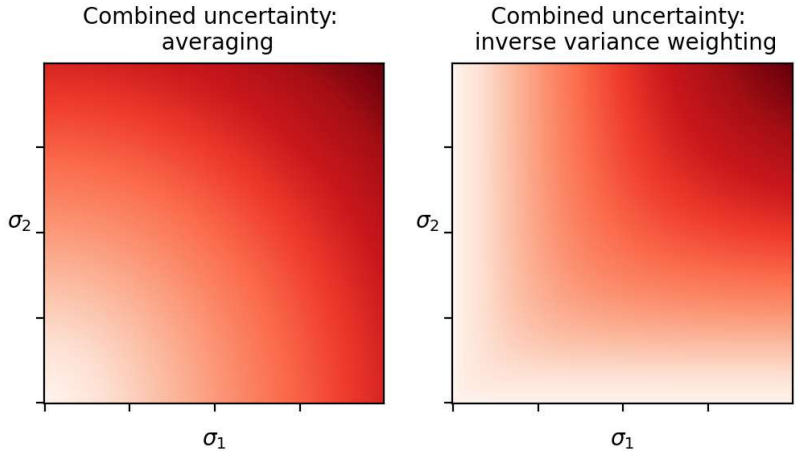


Figure 2.8: Difference in total uncertainty between measurement combination through averaging and through inverse-variance weighting. The two methods are identical along the diagonal.

While useful, the practical applicability of sensor fusion methods such is limited by the ability to perform repeated independent measurements of the desired variable. Correlation between the measurements may cause a bias error in the weighted estimator: measurement correlation prevent an error to average out over repeated measurements because the error is identical over all measurements. This is an insidious property, since mathematically the measurement uncertainty tends towards 0, providing a false sense of security. It the correlations are explicitly known they can be accounted for by advanced Kalman filtering techniques, but this is often difficult information to obtain in practice.

The IEC 61724 standard requires solar irradiance to be measured by multiple pyranometers located at representative locations throughout the site, with the required number of pyranometers increasing with the size of the plant. Sensor fusion methods as above can therefore be applied in theory, but the resulting irradiance estimate should be treated with some caution. First of all, if the pyranometers are positioned as some distance from each other, there may be natural deviation between the irradiance levels at these positions, for instance due to local shading by clouds. Conversely, even though small differences in readings may exists, the instruments will generally experience the same conditions such as ambient temperature and solar zenith angle. This may result in correlations between their measurements errors, especially if the pyranometers are the same model. Hence, the achieved reduction of uncertainty may be less than what theoretical calculations indicate. With this in mind, combining measurements of multiple pyranometers may not necessarily provide better information compared to calculating performance metrics at a sub-system level using single pyranometer readings.

Section 5.3 of the IEC 61724-1 document indicates that subsystem level assessment of performance is preferable compared to combining readings over the plant.

2.3.4 Uncertainty in Decision-Making Processes

Now that we have established a basic understanding on how uncertainty is represented, let us discuss various ways in which it influences decision-making processes. The GUM provides guidelines on how to handle uncertainty in conformity assessment (JCGM, 2012) and we refer to this document for full reference reading .

The overarching concept of this section is the *decision rule*: what is the course of action given a specific set of data? Decision rules can assume various forms and are not always made explicit, although contractual clauses tend to be explicit in their decision rules by virtue of needing to be enforceable. Implicit decision rules are understood in this thesis to be (partly) based on non-quantifiable factors or phrased in ambiguous terms.

Explicit decision rules can generally be characterized as an interval: a decision is defined on whether a specific value exceeds a certain lower bound or falls within certain specifications. For example: suppose a provisional acceptance certificate for a PV power plant is granted if the measured (P50) PR is greater than a certain guaranteed PR_G; the measured PR must fall within the interval $[PR_G, +\infty)$.

The GUM defines two subtly different intervals in this regard: the *tolerance interval* and the *acceptance interval*. The tolerance interval is the interval in which the *true value* of the quantity under consideration must fall, while the *acceptance interval* is the interval in which the *measured value* must fall. So, in the above example the asset owner desires the true PR (henceforth PR_T) of the powerplant to exceed PR_G and has therefore imposed a requirement that the measured PR (henceforth PR_M) must also exceed PR_G (since there is no way of knowing PR_T exactly). Thus, the tolerance interval and acceptance interval are both $[PR_G, +\infty)$. In this example, the asset owner runs the risk of granting an acceptance certificate while $PR_T < PR_G$. This risk is as high as 50% if PR_M is exactly PR_G. Conversely, if PR_M is slightly smaller than PR_G, the EPC will have to pay liquidated damages to the asset owner even though there is a 50% chance that PR_T actually exceeds the guarantee.

The above illustrated false positive and false negatives are called type-1 and type-2 errors in statistics, respectively. See also figure 2.10, taken from the GUM, illustrating this point.

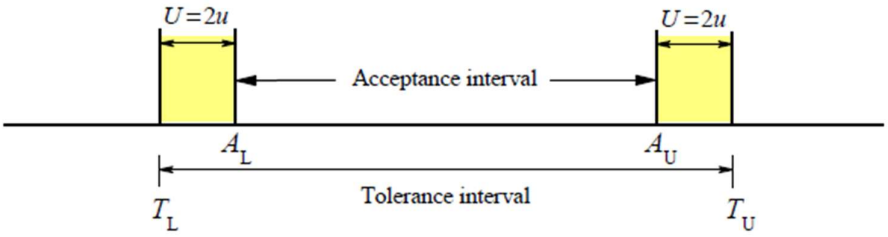


Figure 2.9: Two-sided acceptance interval with guard bands indicated in yellow. The tolerance interval is the interval in which the quantity is desired to be. The guard band is a buffer zone to account for uncertainty in measured values. Only values in the acceptance interval are accepted. Figure taken from the GUM (2012).

There exist various methods that can be employed to mitigate the risk of false positives or negatives. Naturally, reducing measurement uncertainty narrows the potential gap between PR_T and PR_M, which reduces the overall occurrence probability of events where these two values are on different sides of the acceptance criterion. Reducing measurement uncertainty is not always possible, however. Therefore it is common practice to employ guard bands between the tolerance and acceptance interval. *Guard banding* means that the acceptance limit is set inside the interior of the tolerance interval, which reduces the probability of making a type-1 error. A visualization of guard banding as shown in the GUM is provided in figure 2.9.

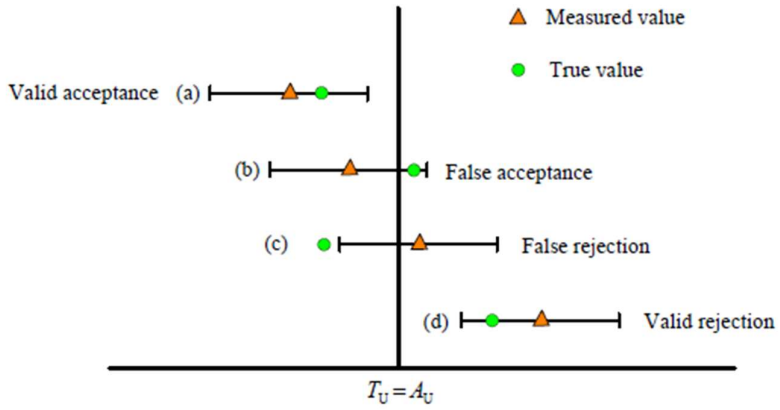


Figure 2.10: Illustration of the 4 cases that can occur during conformity assessment using uncertain data. Figure taken from the GUM (2012).

Guard bands do not need to be explicitly specified. Suppose the asset owner of our previous example desires at least a certain $PR_D < PR_G$, then this PR_D does not need to be included in the contract since only the acceptance interval is of contractual interest. In other words, the uncertainty in the PR measurement is accounted for implicitly by setting a performance guarantee that is higher than the desired performance; guard banding may be in place without this being contractually specified.

What can be seen as a different form of guard banding is changing the metric that is tested, for instance by evaluating the P90 of the PR ($PR_{M,90}$) against the performance guarantee. The decision rule is then whether $PR_{M,90} > PR_G$, which effectively is imposing a guard band equal to the difference between the P90 and the P50. This has the advantage that the risk of making a type-1 error is quantified exactly (10%). However, this requires knowledge of the uncertainty in the PR measurement, something that is not required for regular guard banding.

Be aware that the practice of guard banding is not symmetric in reducing risk for involved parties. Implementing a guard band to reduce the chance of making a type-1 error conversely increases the chance of making a type-2 error.

2.4 Uncertainty in Utility-Scale Solar PV and the Literature Gap

Over the course of this chapter we have discussed the lifecycle of utility-scale solar PV power plants, the role of solar resource measurements during various stages of this lifecycle, PV system performance assessment methods including relevant standards and measurements, and general theoretical concepts of measurement uncertainty. This section goes into greater detail on the uncertainty in the applications of solar irradiance measurements.

For most applications, this section will refrain from going into too much detail, instead we will discuss how uncertainty affects the application and potential challenges with calculating uncertainty. As stated in the introduction, conformity assessment to performance guarantees is the primary lens through which uncertainty will be studied in this research, and will therefore receive a treatment in greater detail.

2.4.1 Inclusion of Solar Resource Uncertainty in the Solar PV Industry

Solar Irradiance Forecasting

Solar irradiance forecasts are combined with computational models of the power plant to obtain energy yield forecasts for the power plant. Uncertainty is an important factor in the uses of energy yield forecasts. For instance, financers of new PV projects usually determine a debt ceiling based on the P75 or P90 of the long-term expected energy production (Pacudan, 2015; IFC, 2015; Mohamadi, 2021; Prilliman et al., 2023). Reducing the uncertainty statistically causes an increase of a project's P75 and P90 energy production forecast, resulting in an increase of the amount of funding that can be raised. While the connection is less direct for short-term forecasts, higher accuracy is, generally, beneficial for the purpose of profit maximization through energy storage and trading algorithms, since the strategy will be optimized on a more realistic forecast. An assessment of the influence of forecast accuracy on revenue optimization strategy has been performed by Campos et al. (2025).

The uncertainty in energy yield forecasts is determined by the uncertainty in the two primary ingredients: the irradiance forecast and the chosen computational model of the PV plant. Advanced software exists that aims to make model versions of solar PV powerplants as realistic as possible. For a review on uncertainty in such models, we refer to Hansen and Martin (2015). A comparison of different available software tools is performed for tropical climates by Raji et al. (2025).

Extensive literature on long-term energy yield assessments and their uncertainty exists. A comprehensive report on this topic was published by the IEA (Reise et al., 2018) which found uncertainties of the order of 2-6% for yearly solar irradiation and 4-8% for monthly solar irradiation. Other studies obtain similar estimates. A report published by the IEA in 2017 lists 3% uncertainty in energy yield (Richter et al., 2017), while a different report published by the IEA in the same year lists uncertainty of 4-8% if the assessment is based exclusively on satellite data and 2.5% if the satellite data is adapted to 3 years of ground measurements.

For short-term energy yield forecasts, there is a wide variety of different methodologies that are employed in different studies. A comparison performed by Di Leo et al. (2025) provides an overview of the different methods.

PV System Performance Assessments

Solar irradiance measurements play a highly important role in PV system performance assessment calculations. We have discussed the measurements and standards involved in this process in section 2.2 and identified multiple sources of uncertainty. Multiple studies have been performed with the aim to quantify the total uncertainty in commonly used performance metrics, most notably the PR. The aforementioned study by the IEA obtained uncertainty estimates for monthly PR of 2-7% (Reise et al., 2018). Other studies arrived at similar or slightly higher estimates: a study by Strobel et al. (2009) reported 2-4.6% uncertainty for annual PR estimates, one by Basson and Pretorius (2016) obtained approximately 2-4% uncertainty for different monthly performance ratios (standard, temperature corrected and weather corrected). A conference proceeding written by Özkalay et al. (2022) lists uncertainty in PR estimates obtained using pyranometer data to be 3.6%, while estimates obtained with satellite data are listed as 3.9% (yearly), 4.7% (monthly) and 8.2% (daily).

Other studies have been dedicated to analyzing the uncertainty in the ASTM E2848 linear regression model. According to the latest update (2023) of the standard itself, typical uncertainties lie between 3.5-7.5% (ASTM, 2018). McCarthy, a large US-based EPC firm, published a conference proceeding for PVPMC in which a rigorous uncertainty calculation is presented (Auslender & Emami, 2025). Instantaneous uncertainty estimates between 1-30% were recovered, with high uncertainty being associated with low-irradiance or unstable conditions. Another study performed an uncertainty analysis of the ASTM metric through Monte Carlo simulations (Newmiller et al., 2025), achieving uncertainties between 3-6%.

In section 2.1.4, various uses of PV performance metrics were discussed, which included plant valuation calculations, informing maintenance decisions, determining degradation rates and as a guarantee in EPC and O&M contracting. We will discuss how uncertainty affects the other use-cases before treating performance guarantees.

Methodologies to determine plant degradation rates over time generally involve some form of statistical inference from a performance metric or a predictive power model timeseries, as discussed in a study by the IEA (French et al., 2021). The linear regression formula of the ASTM E2848 standard is an example of a predictive power model. The publication outlines different metrics to use as a basis and multiple methodologies that can be used to infer degradation rates from these timeseries such as regression and bootstrapping. The study obtained degradation rates for on different datasets ranging between 0.03% to 3.96% per year, with relative uncertainties between 2-12%. The study makes note of considerable dependency of the obtained degradation rate on the chosen model, statistical method and applied data filters.

Plant valuations are based on discounted cash flow models for the remaining lifetime of the power plant, but empirically determined performance helps inform the expected future energy yield that generates this cash flow. Hence, plant valuations are also affected by the uncertainty in long-term energy yield forecasts. To further complicate matters, proper calculation of the future cash flows would also account for degradation using the empirically determined degradation rate. Thus, this would lead to a combined uncertainty in the performance metric, degradation rate and energy yield assessment that affects the projected cash flow that may be difficult to calculate in practice.

Let us now discuss the inclusion of uncertainty in performance guarantees in EPC and O&M contracting with greater detail. EPC and O&M contracts outline post-construction acceptance tests to determine the quality of the completed powerplant, while O&M contracts specify periodic tests of performance metrics against guaranteed values to assess maintenance quality. In a way, these assessments can be seen as the easiest use-case of performance metrics for which we can assess the influence of uncertainty, since we are not using historical data to predict future energy yield or attempting to statistically infer plant degradation from the data, but calculating the historical performance using measured historical data.

Performance guarantee conformity tests are contractually formulated as a comparison between a measured performance metric and a guaranteed value of this metric that should be achieved. In case the performance guarantee is not met, the EPC or O&M contractor is obligated to pay damages to the asset owner. The amount of liquidated damages is often proportional to the absolute deficiency in the performance metric. Some literary sources mention the presence of a minimum PR set below the guaranteed PR in EPC contracts (PWC, 2024; SolarPower Europe, 2021), if the measured PR is below this cutoff the asset owner has the right to reject the powerplant and be reimbursed for all costs. The report by PWC sets this minimum PR as 95% of the guaranteed PR.

Publications by the IFC (2015), the Solar Bankability Consortium (2016), the IEA (Richter at al., 2017; IEA PVPS, 2022) and PWC (2024) frame the conformity assessment as a simple comparison between a measured value and a priorly agreed upon guaranteed value, and a potential triggering of damage claims in case the performance guarantee is not met. In mathematical terms:

$$PR_M \geq PR_{G.} \tag{2.15}$$

This implies either a disregard to the consequences of uncertainty, or the presence of implicit guard bands in the agreed-upon performance guarantee. Given industry awareness of measurement uncertainty the latter possibility appears likely.

A report by the National Renewable Energy Laboratory, or NREL, defined an acceptance test in which a tolerance for errors is included (Dierauf et al., 2013). More explicitly, the plant conforms to the guarantee if the PR is greater than the 95%, or a different contractually agreed upon level, of the guaranteed PR:

$$PR_M \geq (\text{Contract Tolerance})\% \cdot PR_G. \tag{2.16}$$

Observe that the risk is shifted to the asset owner in this case: the effective performance guarantee is lower than the nameplate performance guarantee.

Swami et al. (2024) mention a similar capacity test using a performance guarantee in the ASTM E2848 framework, based on a plant capacity (at reference conditions) P. The measured capacity should be greater than 95-97% of the modeled capacity, where the listed percentages are mentioned as "common" acceptance values. Lacking further elucidation, this can be interpreted as an implicit guard band against false rejections where the asset owner carries the risk. Again:

$$P_M \geq \mathrm{P_G} = (95\ to\ 97)\% \cdot P_{modeled}. \tag{2.17}$$

However, in a separate section this acceptance criterion is rephrased to a rather peculiar statement: the capacity test is passed "if the upper confidence bound of the measured capacity is greater or equal to the guaranteed capacity", which reads in equation form as

$$P_M + U_{95,P_M} \geq P_G. \tag{2.18}$$

In other words, the P2.5 value of the measured energy should exceed the guarantee.

The peculiar nature of this criterion can be intuitively illustrated: suppose the performance criterion in a post-construction acceptance test is a capacity of 10 MW. If the capacity test yielded a capacity of 9 ± 1 MW at k=1.96 the plant passes the test, as well as for 8 ± 2 MW, 6 ± 4 MW, 5 ± 5 MW, et cetera. The increased risk in case of imprecise measurement is shifted wholly to the asset owner, while the EPC or O&M contractor is the party conducting the test; the contractor can lower the bound of the acceptance guarantee by performing lower accuracy measurements.

While not mentioned in the literature, we can construct a reciprocal performance criterion in which the risk for low-accuracy performance measurements is shifted to the contractor, who is performing the test. This would correspond to an acceptance criterion of the form:

$$P_M - U_{95,P_M} \geq P_{\mathrm{modeled}}. \tag{2.19}$$

Criteria at different confidence levels than 2.5% or 97.5% can be readily constructed by multiplying the measurement uncertainty with the k-value of the desired probability level, that can be obtained through the probit function mentioned in section 2.3.2. We can moreover reformulate these criteria using performance ratios instead of plant capacity.

The mention of performance guarantees testing with exceedance probabilities is, with respect to the literature reviewed for this study, only mentioned by Swami et al. (2024). Other sources on this topic either frame the process as a direct comparison like equation (2.15) or a guard banded comparison like equation (2.16), that notably do not require an assessment of the measurement uncertainty to determine.

2.4.2 The Literature Gap

We can divide the function of solar irradiance measurements in two broad categories: as input for solar irradiance forecast calculations and as input for solar PV system performance assessment, that are subsequently used for many different purposes by the solar PV industry. We have seen in literature that solar irradiance measurements are an important source of uncertainty, with stated magnitudes ranging between 2-5% for ground measurements using class A pyranometers (Konings and Habte, 2015). However, studies that calculate the effect of uncertainty in their two use-cases invariably include this as a flat percentage uncertainty, while it was seen in section 2.2.2 that uncertainty of pyranometers is caused by a multitude of sources. Furthermore, these sources are not identical in their contribution to the total uncertainty but have their own characteristics. To the best of the author's knowledge, as of yet no studies have been published in which solar irradiance uncertainty from pyranometer measurements has been divided into its constituent parts before inclusion in subsequent uncertainty analyses of energy yield forecasts or performance metric calculations.

Given the widespread adoption of pyranometers and reliance on their measurements, this motivates a study into the individual contributions of the constituent sources of pyranometer uncertainty and their propagation into energy yield and performance metric uncertainty. Separate inclusion of these sources may enable identification of potential improvement avenues for pyranometers with maximal impact. The 'impact' of a reduction in pyranometer uncertainty can best be assessed through the lens of an industrial application that is influenced by pyranometer uncertainty. To this end, it makes sense to select an application that is least affected by uncontrollable factors, in which the role of pyranometer data is explicit and that lends itself for rigorous uncertainty analysis.

The most suitable application to choose to this end is acceptance testing to performance guarantees. This process involves quite direct comparisons of measured values to guaranteed values with a clear financial interest to the involved parties that can be used as a method to quantify the impact of uncertainty reduction. Furthermore, the influence of uncertainty in this process in a rigorous mathematical way has not been previously performed. Thus, while the use-case lends itself to assess the impact of individual pyranometer uncertainty sources, calculations of uncertainty and financial risk in the acceptance testing process will also provide new insight to the industry in general.

Performance guarantees are typically only part of turnkey projects or projects with external O&M service contractors. Literature states that turnkey projects are the most common type of utility-scale PV development structure (IFC, 2015; PWC, 2024), and that O&M is usually outsourced to third-party providers (NREL, 2018; SolarPower Europe, 2021). Hence, the usage of performance guarantees is not restricted to a niche segment of the market. It should be noted that literature also reports an increasing reluctance among third-party O&M contractors to provide performance guarantees, by virtue of performance being significantly determined by construction quality for which these contractors are not responsible (SolarPower Europe, 2025).

This study is limited to this single application to maintain a focused scope, but that does not imply a study on the influence of pyranometer uncertainty reductions with respect to other applications is not merited. Importantly, studied applications should allow for propagation of pyranometer uncertainty to the uncertainty in the final result. Section 2.4.1 provides reference readings as starting points for potential future research with the aim to explicitly assess the role of pyranometer uncertainty in other applications.

To conclude, this work aims to add to the existing literature in multiple ways. First, by performing uncertainty calculations of the performance ratio with explicit separation of different uncertainty sources, which has not been done previously. The performance ratio is the most-used performance metric and is the easiest to calculate, being independent of computational PV plant models in contrast to the EPI or the ASTM performance evaluation method. These calculations provide novel insight to the contributions of individual uncertainty sources, allows to compare different pyranometers based on their specifications and quantify the impact of uncertainty reductions. Second, through an assessment of the financial consequences of measurement uncertainty in performance guarantee acceptance testing, for which no results are publicly available at present. Together, these two calculations combine to provide a practical quantification of the impact of accuracy improvements in solar resource measurement equipment.

3. Methodology

3.1 The Utility-Scale PV Sector: Literature Analysis & Expert Interviews

The goal of this work was to assess the role of pyranometer measurements in the utility-scale solar PV industry and to investigate the significance of measurement uncertainty in the use cases of pyranometers to inform potential avenues of improvement. The use cases of solar irradiance measurements were identified through a literature study into various interconnected topics. Reviewed literature concerned a wide range of subjects, including utility-scale PV project development, construction and maintenance management, legal and technical best-practices or standards, and scientific literature on the technical state-of-the-art in relevant fields such as yield forecasting and irradiance measurement. An initial literature study informed a model of the utility-scale PV power plant lifecycle, and helped identify the most important use-cases of solar irradiance measurement: plant performance assessment and energy yield assessments in the short- and long-term. It was concluded that the role of measurement uncertainty in performance assessment of PV power plants was the most promising direction to study in this work.

Subsequent desk research was specifically directed at obtaining a deeper understanding of the practical inclusion of irradiance measurements and their uncertainty in performance assessments, and how performance metrics are used as instruments to guide decision-making in industry. Specific attention was given to researching authoritative standards performance assessment methods and relevant measurements, literature discussing relevant legal practices in performance assessment of utility solar power plant construction and maintenance, and to technical studies quantifying uncertainty in pyranometer measurements and performance metric calculations.

The theoretical background discussed in the previous chapter, specifically sections 2.1, 2.2 and 2.4, is a compilation of the relevant literature found during this part of the study, and is the main result of this desk research.

Aside from consulting literature, different qualitative methods were employed to deepen the understanding of the utility-scale solar PV sector and its current challenges. The author attended several meetings of the ongoing revision process of the IEC 61724 standard, during which industry professionals with different backgrounds were invited to share their opinions on the limitations of current performance monitoring practices, and comment on each other's views. Although these meetings were not directly included in this work, some points raised during these meetings helped bring specific topics for further investigation to the author's attention.

Furthermore, two rounds of interviews with industry experts were performed as part of this work. The purpose of this research component was to verify and elaborate upon the process of EPC acceptance testing and the inclusion of uncertainty as described in literature. Due to the limited number of conducted interviews the data is anecdotal and therefore not generalizable to conclusions about industry practice.

The first round of qualitative research took place during the early stages of the research and was aimed at obtaining a general understanding of industry position on uncertainty to help steer the research goals. This stage consisted of unstructured conversations at a solar asset management conference with professionals from various companies. Conversations were held with developers, asset owners, EPC and O&M contractors and SCADA (supervisory control and data analysis) system providers. The interviews were not recorded and no written informed consent was obtained; however, the interviewees were made aware of the purpose of the conversation.

The second round comprised interviews with selected industry experts. These interviews were recorded and written informed consent was obtained from the participants. Because these interviews constituted gathering information from human subjects, this component of the research was subject to university research ethics regulations. Participant informed consent forms and the research data management protocol were prepared in accordance with these regulatory standards, and approval by the TU Delft Human Research Ethics Committee was obtained prior to conducting the interviews.

The interview questions were designed to obtain insight about specific topics regarding inclusion of uncertainty in EPC contracting and the calculation of uncertainty budgets. The interviews were conducted in a semi-structured format, in which the questions served to guide an open-ended discussion rather than being asked in a question-and-answer format.

Since the data of these interviews is not intended to support statistical claims, no structured interview data processing methodology such as response coding was employed.

3.2 Performance Ratio Uncertainty Quantification

3.2.1 GUM Approach

Uncertainty calculations are central to our methodology. It is therefore important that we are rigorous in our application of uncertainty propagation theory throughout our analysis. This work adheres strictly to the Guide to the Expression of Uncertainty in Measurement (the GUM). The theory on expression and usage of uncertainty established in section 2.3 thus applies during this research. In this chapter we will elaborate on the more mathematical aspects of uncertainty and its propagation. Of particular importance are the GUM's guide to the expression of uncertainty in measurement (JCGM, 2008), and the extension of this formalism to any number of output quantities (JCGM, 2011).

The GUM is an easy-to-use approximate method for uncertainty quantification with sufficient accuracy for this thesis. It should be noted that it is not the only available method for uncertainty propagation, analytical calculation of uncertainty or numerical methods can also be used to this end. However, analytical uncertainty evaluation methods are highly nontrivial and beyond the scope of this thesis, and the deterministic nature of the GUM method is preferable over stochastic numerical methods, such as Monte Carlo simulations, for the purposes of this thesis.

The guide to the expression of uncertainty in measurement outlines a procedure of calculating the uncertainty of a quantity Y that is dependent on multiple uncertain input quantities $X_1, \dots, X_N$ through the expression

$$Y = f(X_1, \dots, X_N). \tag{3.1}$$

The uncertainty in Y is determined by three distinct elements: the magnitudes of the uncertainties u_i in each dependency X_i, the *sensitivities* s_i to each of these uncertainties and the correlations $c_{i,j}$ between uncertainties. The total uncertainty u_Y of Y is then calculated as

$$u_Y = \sqrt{\sum_i \sum_{j>i} \left((s_i u_i)^2 + 2 s_i s_j u_i u_j c_{i,j} \right)}. \tag{3.2}$$

We have already discussed the meaning of uncertainty extensively in section 2.3. However, let us devote some attention to the sensitivities and correlations as well.

In the GUM framework, the sensitivities are estimated through a (first-order) Taylor expansion of the function f to the variables X_i,

$$s_i = \frac{\partial \mathrm{f}}{\partial \mathrm{X_i}}, \tag{3.3}$$

that are evaluated at the measured value of X_i. This method is not exact except for functions linearly dependent on X_i. However, the accuracy of this approximation is generally acceptable except for functions that depend on the input quantities in a highly nonlinear way. In such cases the GUM advises to use higher-order approximations or to resort to analytical or numerical methods for uncertainty quantification.

The second important consideration is the correlations between uncertainty sources, also referred to as cross-correlation. In statistics, *correlation* is defined as the extent to which two random variables are linearly related. The concept is closely related to covariance, which is defined as

$$cov(X_i, X_j) = u_i u_j c_{i,j}, \tag{3.4}$$

with $c_{i,j}$ the correlation between the two variables.

Knowledge about the cross-covariance of two random variables is essential for accurate assessment of variable Y's uncertainty, but is often quite difficult to obtain in practice. Generally, there are two methods to obtain this information: analytically or through measurement.

Analytical calculation of the covariance between two error sources requires a deep understanding of the root causes of the uncertainty sources under consideration, on which variables they mutually depend and how this mutual dependency subsequently translates to correlation. Covariance assessment through measurements is typically more practical, but requires strict control of factors potentially influencing measurement. Moreover, the degree of correlation may depend on environmental variables such as ambient temperature, which should be accounted for.

It is generally safe to assume uncertainties of different physical components (such as plant power metering equipment and the pyranometer) are not significantly cross-correlated. However, to the best of the author's knowledge, no cross-covariance characterization has been performed for internal pyranometer uncertainty sources. Performing our own cross-correlation characterization will be a full research unto itself and is outside the scope of an MoT thesis. Therefore we are left to make assumptions on the cross-correlation. Lacking any further information there are two natural options: maximal or minimal correlation. Maximal correlation ($c_{i,j} = \pm 1 \; \forall \; i, j$) means that all uncertainties are fully proportional to each other, while minimal correlation ($c_{i,j} = 0 \; \forall i, j$) all uncertainty sources are fully independent. Since maximally-correlated uncertainty sources are essentially all dependent on a single random variable, the latter option seems overall more physically intuitive. This assumption has also been made by Konings and Habte (2015) in their pyranometer uncertainty quantification.

Hence, in this work we will assume that all internal uncertainties in a pyranometer are mutually independent, i.e. $c_{i,j} = 0 \; \forall i, j$.

The performance ratio is typically not an instantaneous value, but rather a ratio of time-aggregates. What this means for our uncertainty calculation is that we need to assess the uncertainty of a sum of random variables. The GUM provides a guideline for this task in an extension to the main document (JCGM, 2011). The full mathematical derivation of the uncertainty of a time aggregate is presented in appendix A. The result, for independent uncertainty sources, is given by equation (A.28).

Extension to timeseries data presents another dilemma, namely the autocorrelation of uncertainty sources over time. Autocorrelation is similar to the previously discussed cross-correlation, but instead gives the relationship between the instantaneously assumed values of an uncertainty source at two different moments in time. It is given by a coefficient $c_{i;t,\tau}$.

While neglecting cross-correlation terms between sources is an assumption that is physically reasonable, this is not the case for temporal autocorrelation of uncertainty sources. A simple counterexample is an error that is an unknown offset (such as zero-offset A), that are defined by the fact that they remain constant over time and are thus maximally autocorrelated: $c_{i;t,\tau} = 1 \; \forall t, \tau$. However, other uncertainty sources may exhibit noise-like (uncorrelated) behavior over time, or may be correlated up to a certain correlation length (exponential decay of correlation). On top of this, uncertainties may consist of both an offset and a randomly fluctuating part, although the offset is usually minimized through calibration procedures. Analytical quantification is difficult and experimental autocorrelation characterizations have not been performed. This study will assume full autocorrelation of uncertainty sources over time.

An advantage of working with independent uncertainty sources is that it allows for a direct quantification of the sensitivity to the magnitude of a single uncertainty source. From the mathematical analysis in appendix section A.4 we see a quadratic sensitivity of the total uncertainty to individual uncertainty sources, with the scale determined by the relative contribution of this source to the total uncertainty budget (equation A.33). This means that, under the assumption of independent uncertainty sources, reducing the uncertainty of the most significantly contributing uncertainty source has a proportionally more significant impact in reducing the total uncertainty compared to reducing the uncertainty of sources with a smaller relative contribution.

3.2.2 Included Uncertainties

The uncertainties and the magnitudes of the uncertainties included in this study will be taken mostly from the various standards discussed in the previous chapter.

- For pyranometers, we will assume ISO 9060 class A uncertainty sources and bounds and expand upon these by also including datalogger accuracy. The uncertainty estimate for datalogger accuracy is taken from Konings and Habte (2015).
- Pyranometer calibration uncertainty is taken from the ISO 9847:2023 estimate for class A pyranometers.
- Pyranometer non-stability (or calibration drift) uncertainty is included with the assumption that the last calibration was performed one year ago. The IEC 61724-1 document prescribed calibration at least every two years, the included uncertainty thus amounts to an 'average expected' calibration drift.
- Uncertainty of the temperature and power measurements will be within the IEC 61724 specifications of these measurements.
- Module calibration uncertainty will be estimated from literature. In this study, the values from Dirnberger and Kraling (2013) will be taken.
- Uncertainty due to soiling will not be included.

The purpose of the chosen uncertainty bounds is to obtain a realistic estimate of achievable uncertainty of state-of-the-art monitoring systems, hence class A specifications are followed wherever possible. Pyranometer calibration is recommended once every two years by manufacturers (Hukx, n.d.-a), hence we will include pyranometer non-stability assuming the last calibration was performed 1 year ago. Moreover, we assume the pyranometer is regularly cleaned, so uncertainty due to soiling is not included.

Since the datasets used for this work do not include direct solar irradiance measurements we must make an approximation for the directional response uncertainty bound. We employ the recommendation by Konings and Habte (2015) to calculate this bound with respect to the measured plane-of-array solar irradiance instead.

A summary of all included uncertainties and their magnitudes is listed below. As motivated in section 2.2.2, soiling will be excluded from this work due to its highly variable nature. The influence of soiling in different conditions and under different cleaning regimes merits a dedicated methodological study.

Table 3-1: Uncertainty bounds used for uncertainty calculations.

Uncertainty Source	**Acting on**	**Error bound**	**Specifications**
Zero off-set A	G_{POA}	$7\ W \cdot m^{-2}$	Rectangular, Type B, one-sided
Zero off-set B	G_{POA}	$2\ W \cdot m^{-2}$	Rectangular, Type B
Zero off-set other sources	G_{POA}	$1\ W \cdot m^{-2}$	Rectangular, Type B

Additional signal processing errors	G_{POA}	$2\ W \cdot m^{-2}$	Rectangular, Type B
Directional response	G_{POA}	$10 \frac{G_{POA}}{G_{STC}\ cos(\theta)}\ W \cdot m^{-2}$	Rectangular, Type B
Calibration	S	1.0 %	Rectangular, Type B
Non-stability	S	0.8 %	Rectangular, Type B
Nonlinearity	S	0.5 %	Rectangular, Type B
Clear-sky spectral error	S	0.5 %	Rectangular, Type B
Tilt response	S	0.5 %	Rectangular, Type B
Temperature response	S	1 %	Rectangular, Type B
Datalogger accuracy	V	$10\ \mu V$	Rectangular, Type B
Power metering	P_{out}	0.2 %	Rectangular, Type B
Back-of-module temperature	T_{mod}	$1\ °C$	Rectangular, Type B
Single module rating at STC	P_0	2.4% ($\sigma = 0.8$ %)	Normal, Type B

The above table lists ISO 9060 class A pyranometer uncertainty specifications. Comparisons are performed with performance ratio calculations using ISO 9060 class B uncertainty specifications, while keeping the other uncertainty bounds constant. The ISO 9060 specifications for class B pyranometers are listed in table 3-2.

Table 3-2: ISO 9060 class B pyranometer uncertainty bounds.

Zero off-set A	$15\ W \cdot m^{-2}$
Zero off-set B	$4\ W \cdot m^{-2}$
Zero off-set other sources	$2\ W \cdot m^{-2}$
Additional signal processing errors	$5\ W \cdot m^{-2}$
Directional response	$20 \frac{G_{POA}}{G_{STC}\ cos(\theta)}\ W \cdot m^{-2}$
Non-stability	1.5 %
Nonlinearity	1 %
Clear-sky spectral error	1 %
Tilt response	2 %
Temperature response	2 %

Comparisons will also be performed with manufacturer uncertainty limit specifications of industrial state-of-the-art pyranometers, since commercially available class A pyranometers are often more accurate than prescribed in the ISO 9060 standard. In this work, the Hukx SR300-D1 pyranometer will be used to compare the accuracy of market-leading pyranometers with ISO requirements. The uncertainty limits of the SR300-D1 pyranometer as listed in the product user manual (Hukx, n.d.-b) are shown in table 3-3.

Table 3-3: Hukx SR300-D1 industrial class A pyranometer uncertainty specifications.

Zero off-set A	$2\ W \cdot m^{-2}$
Zero off-set B	$2\ W \cdot m^{-2}$
Zero off-set other sources	$1\ W \cdot m^{-2}$
Additional signal processing errors	$0\ W \cdot m^{-2}$
Directional response	$10 \frac{G_{POA}}{G_{STC}\ cos(\theta)}\ W \cdot m^{-2}$
Calibration	0.6 %
Non-stability	0.5 %
Nonlinearity	0.2 %
Clear-sky spectral error	0.5 %
Tilt response	0.2 %
Temperature response	0.4 %

3.2.3 Calculation Methodology

The uncertainties listed in tables 3-1, 3-2 and 3-3 are not expressed in the same units. Some uncertainties are specified in relative terms (percentages) while others are listed as absolute uncertainty limits. A purely theoretical calculation of the total uncertainty requires all magnitudes to be specified in relative terms and is therefore not possible in our case. Consequently, empirical data from solar PV systems must be used to convert all uncertainty components to the same units. We will address the datasets used for this purpose in section 3.2.4.

Previous studies have been performed in which pyranometer uncertainty is calculated by decomposing it to the root sources. The aforementioned ASTM G213-17 (2017) standard guide provides users with an excel spreadsheet that can be used for these calculations. The spreadsheet allows to calculate the instantaneous and aggregated uncertainties for a day of 1-minute pyranometer data. The spreadsheet also already includes all pyranometer uncertainty sources listed in table 2-1. This makes it a natural starting point for this study. If the spreadsheet can be extended to include power and temperature data and their uncertainty it can be used for performance ratio uncertainty assessment with explicit inclusion of pyranometer uncertainty sources.

There are several complications with this approach. To start, excel is not suited for the large matrix calculations that arise from aggregating temporally autocorrelated uncertainty (a square matrix of size 1446 for a single day of 1-minute data), which has lead the creators to improvise a method to do these calculations step-wise in excel by first aggregating to hourly and subsequently to daily uncertainties. Moreover, the various uncertainty sources are aggregated first and the total uncertainty is temporally aggregated second. Time-aggregation of temporally autocorrelated uncertainty is not an associative operation with itself, nor is aggregation of independent autocorrelated uncertainty sources associative with time-aggregation. Hence, this methodology leads to mathematically incorrect results. A formal derivation of this error is presented in appendix A.

Additionally, excel is ill-suited for this operation from a flexibility perspective. Extension to more uncertainty sources requires either making cumbersome expansions of existing sheets, or manual copying of existing uncertainty calculation procedures for new uncertainty sources in new sheets. Moreover, excel spreadsheets do not allow for easy parsing through large amounts of data or preprocessing this data.

Because of the calculation errors in the ASTM G213-17 spreadsheet and the inextensible nature of excel spreadsheets as data analysis method, the uncertainty calculations in this work will be performed using custom python software.

The python software used for the uncertainty calculations was specifically developed for the purposes of this thesis. Priorities in the design for the software were flexibility with respect to the chosen performance metric, the included uncertainty sources and data preprocessing methods. The code is publicly available through github and a full documentation of the code is provided in appendix C. A link to the repository is also provided in the appendix.

The code's intended functionality is to propagate uncertainty through a *tree* or an *acyclic directed graph* of variables, that are related to each other through equations. We will refer to this as the *equation tree* going forward. An example of the equation tree for the temperature-corrected performance ratio is shown in figure 3.1. Each variable in the equation tree can be given any number of uncertainty sources acting on it. These uncertainty sources can have their magnitudes, distributions and autocorrelations individually specified. Importantly, all uncertainty sources are assumed to be independent.

Each variable in the equation tree is either pre-defined (constants), read from a CSV file (time-series of measurements) or calculated from other variables through a defined equation. Upon execution, the code will load an equation tree, checks whether it is well-defined, populates variables from specified CSVs, performs optional data preprocessing and calculates all values and uncertainties of the variables in the equation tree.

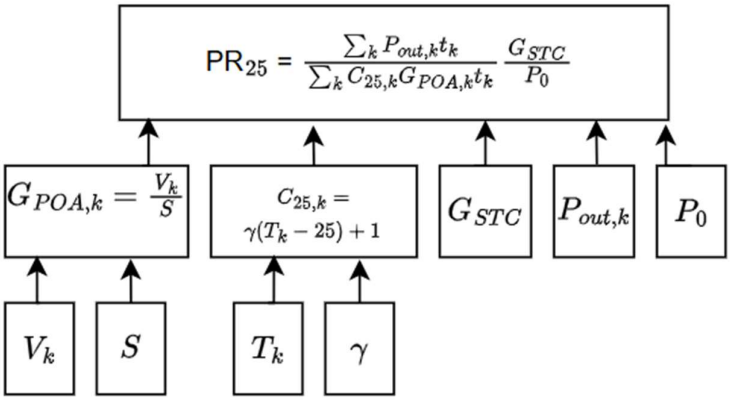


Figure 3.1: Illustration of the equation tree for the temperature-corrected performance ratio.

The code is designed such that the procedure outlined in the previous paragraph can be easily iterated over for multiple CSVs. Additionally, automatic time-matching of CSVs with different temporal resolution can be performed.

Data preprocessing steps can be easily integrated in the process. Data preprocessing is used here as an umbrella term investigating the presence of missing or unphysical datapoints. Unphysical datapoints can be widely interpreted, it may for instance include nonzero irradiance measurements during nighttime or measured values above or below a maximum allowed value. The code integrates the pvlib python library to optionally calculate the solar zenith angle at each datapoint. The solar zenith angle tells us whether we expect irradiance and power output measurements to be zero (for angles greater than 90°) or non-zero (for angles smaller than 90°). Preprocessing functionalities also include imputation of missing data through linear interpolation that can be performed if desired.

3.2.4 Data for Uncertainty Calculations

This research has employed two different datasets containing solar irradiation, power output and temperature data spanning 1 year. The purpose of this research is not to quantify the uncertainty budget of a single utility-scale power plant as closely as possible. The goal is rather to obtain the theoretical uncertainty budget of a powerplant when adhering to uncertainty specifications of widely used standards, and from there to identify promising avenues of uncertainty reduction and the theoretical impact of these reductions in acceptance testing.

Therefore, the data served to facilitate this assessment by allowing us to convert relative bounds to absolute values and vice versa. While datasets of utility-scale power plants are preferable, these are often proprietary and not publicly available. Data of small-scale setups was therefore used in this research.

Furthermore, datasets often exclusively report the measured POA irradiance instead of the analogue pyranometer output and a sensitivity. In order to properly calculate all uncertainties, the output voltages will be determined from the listed irradiance using an assumed sensitivity of 15 $\mu V \cdot W^{-1}m^2$, which was also used by Konings and Habte (2015). Similarly, datasets will not always include back-of-module temperature. The back-of-module temperature will in such cases be estimated through formula (2.10), as provided in the IEC 61724-2 document. The formula is repeated here for convenience:

$$T_{\mathrm{mod}} \approx G_{POA} \cdot \left(e^{a+b \cdot v_W}\right)[^{\circ}Cm^2/W] + T_a. \tag{3.5}$$

The estimated module temperature will be treated as if it was measured using a temperature sensor. The intended application of this work's results is for utility-scale power plants, which have dedicated back-of-module temperature sensors. Including the epistemic uncertainty of the above formula in the analysis is therefore not relevant.

The following datasets were selected:

- Dataset 1 is from a research setup on the roof of the SolarTech lab of the Polytechnic University of Milan, located at (N 45° 30'10.3'', E 9° 09' 23.66'') and was made available following a publication by Leva et al. (2020). The dataset contains 1 year of power output, POA and global horizontal solar irradiance, ambient temperature, wind speed and humidity data at single-minute resolution. The data covers the period from 1-1-2017 to 31-12-2017. However, in the accompanying paper it is stated that not all days are available due to faults and disconnections.
  The experimental setup consists of a single (monocrystalline) PV module tilted at 30° with a listed power rating of 245 W_p. It has a listed measurement accuracy for the output power of $\pm 3\%$, but this will be replaced by the IEC 61724-1 bound for class A monitoring systems. This is listed as 0.2% for regular operating conditions (unity power factor and operating current greater than 5% of the maximum rated current). The used pyranometer is listed as ISO 9060 class A compliant. The maximum-power temperature coefficient is assumed to be $\gamma = -0.45\%$ for this module, which is taken from the datasheet of a PV module of similar specifications (Suntech STP245S – 20Wd module; monocrystalline PV module rated at 245 W_p).
- Dataset 2 is from a ground-mounted PV powerplant near Irece, Brazil, located at (S 11° 19' 35", W 41° 51' 56"). The dataset was made available accompanying a publication by Ribeiro et al. (2021). The datasets contains 1 year of power output, voltage, current, module temperature, ambient temperature, POA irradiance, humidity, wind speed and other environmental data, spanning from 1-8-2018 to 31-7-2019. The output power and module temperature data is of 15-minute resolution while the environmental data is of 1-minute resolution. The accompanying publication states missing data during May and early June due to an inverter malfunction. Data inspection showed this malfunction lasted from April 17th to June 12th.
  The data concerns a single array totaling 96 PV modules tilted at 15°. The modules are rated at 97.5 W_p with a maximum-power temperature coefficient $\gamma = -0.29\%$.

Data will be separated by day and the PR and temperature-corrected PR and their uncertainties will be calculated separately for each day. Before calculation, the single-day data will pass through a preprocessing step which does the following:

- Check for unphysical and extreme values and change these to 'missing data points'. These include: negative power or irradiance data and temperatures outside environmentally expected ranges (-10 to 50 °C).
- Interpolate single missing values: if a missing data point is bordered on both sides by regular data, the missing datapoint will be linearly interpolated by these values.
- Check for remaining missing data during daytime hours. This can be done either by comparing to the solar zenith angle or by checking whether the missing values are contained in the 'body' of the data.

- Check for non-zero data during nighttime hours for data expected to be zero at this time. Done by comparing to the solar zenith angle. Optional cleaning of this data in case the errors are small.

The data preprocessing steps are designed to filter out incomplete days, while being flexible enough that isolated missing datapoints will not invalidate an entire day of data.

After computation a postprocessing step is performed that automatically flags days where calculated quantities such as the PR assume uncommon values, prompting manual inspection of the cause of this data. It may be the case that days containing faulty data passed through the preprocessing filters.

A meaningful extension to this research would be to repeat the uncertainty calculations using more datasets that preferably concern large to utility-scale PV systems. A potential starting point to this end is the publicly available database[3] of PV system data that accompanies the publication by French et al. (2021). The code developed for this research can be applied for this purpose and is publicly available through the repository listed in appendix C.

3.3 Conformity Assessments using Uncertain Information

The literature study revealed that performance metrics play an important role in plant construction and maintenance quality guarantees, giving rise to situations where significant financial interests of asset owners and the contractors for EPC and O&M work depend on the measured value of a given metric. The review in section 2.4.1 identifies various conformity assessment methods for performance guarantee clauses described literature. According to most reviewed sources, uncertainty is usually disregarded completely, or included through an implicit guard band of 3-5%. A single source stated inclusion of uncertainty through by defining an exceedance probability level as the acceptance criterion.

Post-construction acceptance testing is an especially prominent use case of performance metrics. In our review of EPC contracting practices in the utility-scale PV industry, we saw that 5-15% of EPC contract value is often tied to the plant passing the final acceptance test. In case the guaranteed performance is not achieved by the plant, damage clauses usually specify the EPC contractor owes damages to the plant owner that are determined as the net present value of the missed lifetime energy yield of the plant. Moreover, the EPC contractor may be liable to pay the full plant construction cost in case a minimum performance (lower than the guaranteed performance) is not achieved.

Similar clauses are defined for service quality assessment in O&M contracts, which usually define yearly conformity assessments to performance guarantees. Again, the liquidated damages owed to the asset owner in case of underperformance tend to cover the foregone revenue of the missed energy yield, or are otherwise specified as a fixed percentage of the contract value per percentage-point underperformance. Liquidated damages are typically capped at 100% of the O&M contract value, with a clause for possible termination of the contract in case of severe underperformance.

Unique insights may be obtained by placing the calculations of achievable uncertainty in PR (and the relative contributions of the constituent sources) outlined in the previous section in the context of performance guarantee conformity assessment. Specifically, we are interested in determining the financial risk assumed by asset owners and EPC and O&M contractors in current practices, given the calculated magnitude of the uncertainty in performance ratio.

In this work, we define financial risk as the cost incurred by a party in case an event occurs, times the occurrence probability of this event. Mathematically:

[3] Database can be found at: https://osf.io/vtr2s/overview

$$\text{risk} = \$(\text{cost of event}) \cdot P(\text{event}). \tag{3.6}$$

This section will exclude the eventuality that the entire plant is rejected from the consideration, since the cost of such an event differs from standard liquidated damages by two orders of magnitude.

In literature, it is stated that the damages owed by the EPC contractor are equal to the net present value of the missed income over plant lifetime due to the lower-than-guaranteed performance. For O&M contracting, literature stated that the penalty should cover the missed income due to underperformance during the considered year. Since we have seen that conformity assessments sometimes apply guard bands or other adjustments to the guaranteed PR to set a PR at which damage clauses trigger, we will define the latter as the *required* PR. We denote the required and measured performance ratios as PR_R and PR_M.

Determining the liquidated damages is easier in case of underperformance in O&M contracting, since we are dealing with measured values instead of lifetime projected values. The O&M best practices handbook by NREL et al. (2018) provides a calculation of the form

$$\text{Liquidated Damages} = (PR_R - PR_M)_{PR_M < PR_R} \cdot P_0 \cdot p \cdot \text{measured irradiation in kWh}, \tag{3.7}$$

With p the average PPA tariff per MWh of generated energy over the considered period. Typically, the PR_R is lowered each year to account for the projected plant degradation over time.

While the missed revenue relatively straightforward to determine for O&M contracting, this is not so for post-construction acceptance testing, which requires an ex ante calculation of the value of the missed yield over the entire plant lifetime. However, we will apply the same philosophy as in equation (3.7).

The performance ratio is the realized energy yield $\sum_k P_{out,k}\tau_k$, divided by the rated yield of the plant given the measured total hours of STC irradiation: $P_0 \cdot \frac{\sum_k G_k \tau_k}{G_{STC}}$. Correction factors simply append this ratio to account for external factors influencing module or plant efficiency, such as temperature. The performance ratio can thus be used to obtain an estimate of the energy yield over a period of time, given an estimate of the expected solar irradiation during this timeframe (in terms of total hours of STC irradiance, or kWh):

$$\text{Expected yield} = P_0 \cdot PR \cdot \text{expected irradiation in kWh}. \tag{3.8}$$

Naturally, PR is replaced by PR_{25} depending on contractual specifications. Further corrections can be applied to the expected yield to account for downtime or degradation by simply multiplying the rated capacity with their respective loss factors. So, if degradation is 5% we can multiply P_0 by 0.95.

In the event that the measured performance ratio is lower than the required value, we can calculate the missed yield over the plant lifetime according to formula (3.8) as

$$\text{Missed yield} = (PR_R - PR_M)_{PR_M < PR_R} \cdot P_0 \cdot \text{expected irradiation in kWh}. \tag{3.9}$$

Again, extension to include degradation and downtime is straightforward. The obtained expression is also found in literature: the report by SolarPower Europe (2021) presents an identical expression for the foregone yield.

The revenue shortfall is obtained by multiplying the foregone yield in MWh with the price of one MWh of electricity. However, since PV plant lifetime is typically upwards of 20 years, future cash flows must be discounted to their net present value.

Moreover, the price per MWh is also not necessarily constant. We have seen in our treatment of PPAs that physical PPAs often contain escalator clauses, and that virtual PPAs allow sale of electricity against market price.

To keep calculations manageable we view the missed income as an annuity; it is treated as a series of yearly missed incomes. Let us index the years by n, assume a constant discount rate d, the total years of plant lifetime as N, assume a constant linear annual degradation at rate D, a guaranteed plant availability of T_G, and define the price of electricity in year n as p_n. The net present value of the missed revenue, and thus the amount owed to the asset owner by the EPC contractor, is then

$$NPV(\text{missed yield}) = \sum_{n=1}^{N} \frac{p_n}{(1+d)^n} (PR_R - PR_M)_{PR_M<PR_R} \cdot P_0 \cdot (1-nD) \cdot T_G \cdot \text{expected annual irradiation in kWh.} \tag{3.10}$$

For brevity, we define

$$\xi = \sum_{n=1}^{N} \frac{p_n}{(1+d)^n} \cdot P_0 \cdot (1-nD) \cdot T_G \cdot \text{expected annual irradiation in kWh,} \tag{3.11}$$

which allows to write the above expression as

$$NPV(\text{missed yield}) = (PR_R - PR_M)_{PR_M<PR_R} \cdot \xi. \tag{3.12}$$

Observe that equation (3.7) can be written in an identical way by replacing

$$\xi = P_0 \cdot p \cdot \text{measured irradiation in kWh.} \tag{3.13}$$

We see that conformity assessment to performance guarantees is separable into a difference between the measured PR and a required PR (which is determined by the specific acceptance testing clause) and a parameter ξ characterizing the amount of liquidated damages per absolute percentage underperformance.

On the errors in determining amounts payable by the EPC and O&M contractors.

Although the contractor is beholden to explicit financial costs given by the above expression in case of (measured) underperformance, there are also unknowable costs associated with the acceptance testing process. These costs are unknowable because they are caused by imperfect measurement of the unknowable true value of the performance ratio, PR_T.

Firstly, the asset owner may incur financial losses in case a type-1 error (false positive) is made during the conformity assessment. This means that, due to a measurement error, the measured PR_M is greater than the required PR_R, while the unknown true PR_T is actually lower than the required value. Thus, type-1 errors are at the financial risk of the asset owner. In this case, the asset owner will mistakenly accept the plant and forego the damages equal to

$$NPV(\text{missed yield}) = (PR_R - PR_T) \cdot \xi. \tag{3.14}$$

Secondly, in case a type-2 error is made, the contractor will pay the previously derived contractually agreed upon damages to the asset owner, while in actuality the plant conforms to the performance guarantee.

Lastly, there is the case where $PR_T < PR_R$, but the amount of damages deviates from the true value:

$$NPV(\text{erroneous damages}) = (PR_M - PR_T) \cdot \xi. \tag{3.15}$$

A positive value indicates the contractor paid too little, a negative value indicates the contractor paid too much. Thus, these errors can be at the risk of both the asset owner and the EPC or O&M contractor.

To assess the financial risk of the EPC and O&M contractors, we will calculate the expected value of the owed damages. Since expressions (3.7) and (3.10) are identical up to the definition of ξ, the derivation in this section is applicable to both EPC and O&M performance guarantees.

Mathematically, we are interested in the expected value of a function $f(x)$ that is dependent on a random variable x. Moreover, we have the conditional probability distribution of x given the value of a quantity a that is fixed and known. The expected value of $f(x)$ is then given by

$$\mathrm{E}(f(x)|a) = \int f(x)\mathrm{P}(x|a)dx\,. \tag{3.16}$$

In our case, the function f is the cost function of the EPC or O&M contractor and x is the measured PR_M. The conditional variables are the variables which characterize the probability distribution: the unknowable true performance ratio PR_T and the measurement uncertainty u_{PR}.

The probability that a powerplant fails to meet the required performance upon measurement, meaning that the contractor has to pay some amount of damages, is given by

$$\mathrm{P}(\text{fail} \mid PR_T, u_{PR}) = \int_0^{PR_R} \mathrm{N}(PR_M;\ \mu = PR_T, \sigma = u_{PR})dPR_M\,, \tag{3.17}$$

where $\mathrm{N}(x; \mu, \sigma)$ denotes the probability density function of a normal distribution,

$$\mathrm{N}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{3.18}$$

The expected financial cost for the contractor is then, using expression (3.12), given as

$$\mathrm{E}(\text{payment} \mid PR_T, u_{PR}) = \int_0^{PR_R} (PR_R - PR_M) \cdot \xi \cdot \mathrm{N}(PR_M;\ \mu = PR_T, \sigma = u_{PR})\ dPR_M\,. \tag{3.19}$$

We assumed here that the probability density function of the measured PR is normally distributed around the true PR with the measurement uncertainty as standard deviation. This is in accordance with the GUM framework; an (analytical) analysis of the probability distribution of the measured PR is beyond the scope of this work. Realistically, the probability distribution may be skewed or contain a bias to one side.

Another financial risk measure is the α-% value-at-risk, which is the financial loss that is exceeded in the worst α% of cases, with common values for α ranging between 2.5-10%. Mathematically, it is given as

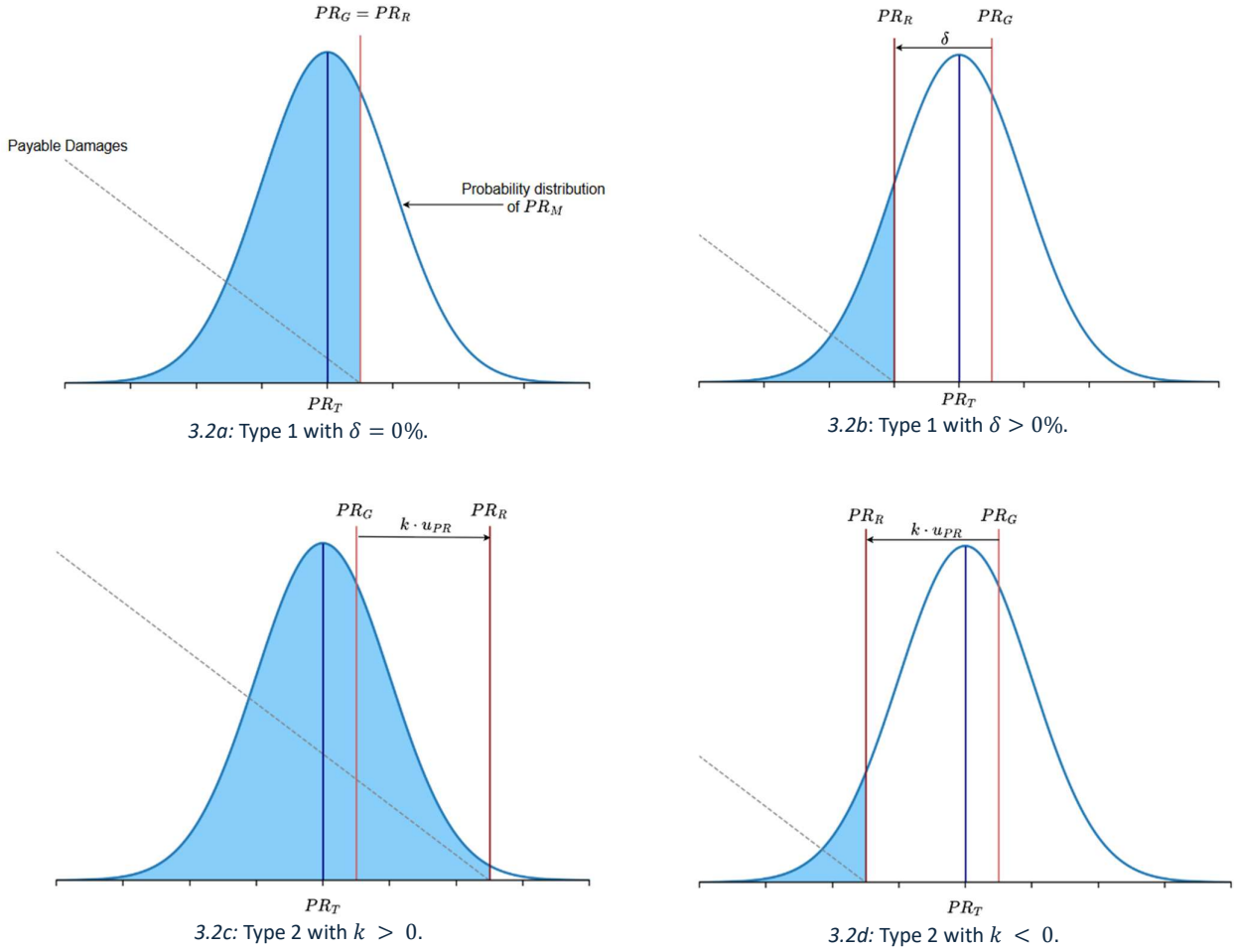
$$\alpha\%\ \text{Value-at-Risk} = (PR_R - PR_M) \cdot \xi \mid_{PR_M = \alpha\%\ \text{quantile of}\ \mathrm{N}(\mu = PR_T, \sigma = u_{PR})} \tag{3.20}$$

The value-at-risk provides insight in the loss that is incurred in worst-case scenarios for the EPC and O&M contractors and thus helps us investigate what happens to the tail-end of the liquidated damages for various cases.

We can distinguish two types of acceptance testing clauses from the literature, either implementing an implicit guard band δ or explicitly including the measurement uncertainty. In this work, the following acceptance testing clauses are considered.

- Type 1: $PR_R = PR_G - \delta$, with $\delta \in \{0\%,\ 3\%,\ 5\%\}$,
- Type 2: $PR_R = PR_G + k \cdot u_{PR}$, with $k = \pm 1, 1.28, 1.96$.

Case 1 is a simple test in which uncertainty is either ignored completely ($\delta = 0$) or is considered through an implicit guard band ($\delta > 0$). In the second case, the conformity assessment is explicitly dependent on the measurement uncertainty. The positive coverage factors yield the 84%, 90% and 97.5% exceedance levels, negative coverage factors their reciprocals.



3.2a: Type 1 with $\delta = 0\%$.

3.2b: Type 1 with $\delta > 0\%$.

3.2c: Type 2 with $k > 0$.

3.2d: Type 2 with $k < 0$.

Figure 3.2: Illustration of different types of conformity assessment following the model in this thesis. Figures show the probability distribution of PR_M for a true PR lower than the guaranteed PR in different testing cases. The shaded area equals the probability that the plant gets rejected ($PR_M < PR_R$). Liquidated damages are shown in the dashed line.

The expected liquidated damages are dependent on three free variables: PR_G, PR_T and u_{PR}, and the testing type. However, it can be characterized by just two degrees of freedom: the uncertainty u_{PR} and the distance between PR_G and PR_T. This is because performance guarantees are usually far higher than 0 and the measurement uncertainties σ (at k=1) are in the range of 1-5%. The probability density of PR_M near 0 is therefore negligible in practical cases, meaning we can safely disregard the degree of freedom in the distance between PR_G and 0.

We assess the risk for the EPC and O&M contractors by determining the expected value of the owed damages, the probability of failing the acceptance test, the error in the owed damages and the 5% value-at-risk for different situations, defined by the acceptance testing clause, the true value of the PR and the value of the PR uncertainty. Because Gaussian integrals over a generic range as above are not analytically integrable these calculations are performed numerically using python. For all 4 acceptance clauses, the probability of failing the test and the expected loss (in units of ξ) are determined for a grid of possible values (u_{PR} , $PR_T - PR_G$).

Importantly, these calculations will disregard the possibility of the power plant being rejected completely in cases of significant underperformance. Little information on such procedures was found in literature and the financial risk for the EPC and O&M contractors associated with this eventuality is orders of magnitude greater than the liquidated damages for slight underperformance. Hence, assessment of this risk is best assessed in the form of occurrence probability. The results of equation (3.17) can be reinterpreted to this end.

In addition to this, an analysis of the relative magnitude of ξ to the total contract value in post-construction acceptance testing is performed using realistic values from literature. Specifically, we use the data from the *2025 utility-scale solar data update* by Seel et al. (2025), which contains recent estimates regarding EPC CapEx, PPA pricing, degradation rates and annual solar irradiation for the US utility PV sector. Values used in this work are the average, or 'P50', values for these variables as listed in this mentioned work. The report does not provide discount rates or PPA escalation factors and has instead performed a discount calculation for the PPA tariff already; the reported PPA prices are the levelized PPA tariffs over plant lifetime. Interestingly, the average levelized cost of energy (total lifetime costs divided by total lifetime energy generation) is reported to be higher than the average levelized PPA price, indicating an average net financial loss over plant lifetime at present.

Table 3-4: Values used for the net present value calculation of the foregone energy in case of plant underperformance. Values are US-industry averages taken from Seel et al. (2025), unless specified otherwise.

Variable	Value	
Levelized PPA tariff	29	\$/MWh
CapEx per Watt capacity	1.61	$\$/W_{AC}$
Degradation rate D	-1.6	% / year
Average STC irradiation	3 / 4.5 / 6	$kWh/m^2/day$*
Guaranteed uptime	99	%**
Operational plant lifetime	23	years***
Power rating	100	MW_{AC}***
Not used for the calculations		
Levelized OpEx	5	\$/MWh
Levelized cost of energy (LCOE)	60	\$/MWh
LCOE including tax credits	41	\$/MWh

*Low, medium and high irradiance values, realistic estimates for different areas of the US, taken from Seel et al. (2025).

** Value taken from SolarBankability Project (2016).

***Realistic values chosen by the author.

4. Results

4.1 Sector Analysis & Interviews

An analysis of the utility-scale solar PV sector was performed in order to understand when and how solar irradiance measurements play are utilized throughout the value chain. Initial desk research focused on mapping the typical lifecycle of utility-scale PV power plants to facilitate identification of different use cases of solar irradiance measurements. This work resulted in the lifecycle model discussed in section 2.1.1 and shown in figure 2.1. The uses of solar irradiance measurements were identified for each part of the lifecycle and can be broadly categorized between energy yield forecasting and PV plant performance assessment. In figure 4.1 the uses of solar irradiance measurements are placed in their respective phases of the PV power plant lifecycle.

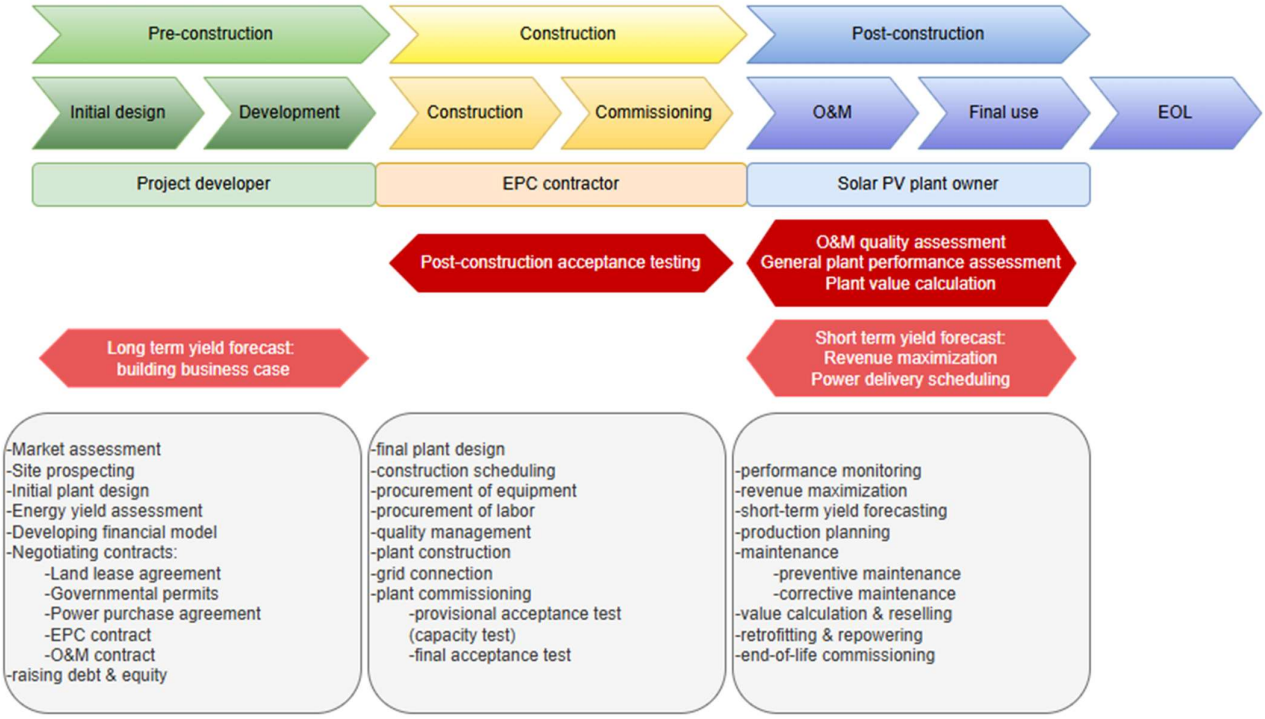


Figure 4.1: Solar PV power plant lifecycle of with the uses of solar irradiance measurements shown per phase of the lifecycle. Upper uses (dark red) concern plant performance metrics, lower (light red) uses are applications of energy yield forecasts.

The purposes and stakeholders were mapped for each use case and technical literature was consulted to understand contemporary methods of calculating energy yield forecasts and performance metrics, with specific attention to the ways in which uncertainty is incorporated. From this analysis and the results of the first round of interviews, the present literature gaps with respect to inclusion of pyranometer uncertainty were identified, with the most promising research avenue being identified as performance metrics and their use in acceptance testing. In literature, performance ratio guarantees in construction and maintenance quality assessment were found to be mentioned by multiple sources as important elements of EPC and O&M contracts, that are accompanied by penalty payment clauses that come into effect when these guarantees are not achieved. Liquidated damages were found to be defined as being proportional to the value of the missed energy yield due to underperformance. As will be discussed in greater detail in section 4.3, such damages can be in the range of millions of dollars for large powerplants. Additionally, uncertainty estimates of performance metrics were found to range between 2-8% in technical literature.

Notably, no work was found in which an analysis was made on the influence of constituent uncertainty sources to this total, the impact of potential improvements to this uncertainty or the influence of uncertainty in the use cases of performance metrics.

From the literature, it became apparent that the post-construction acceptance process of a power plant does not assume one singular form, but varies on a case-to-case basis. Compiling information on this process revealed numerous possible variations, including in the total duration of the test, the requirements on solar irradiance for days to be included in this test, the chosen performance metric and the conformity assessment of the measured performance metric to the guaranteed value. Test duration can range from a few days of high irradiance to all days over the course of two years. The performance metric chosen for the test can be the total plant capacity, or a dedicated performance metric such as the performance ratio. Moreover, the ways in which uncertainty is included in conformity assessment and the associated risk allocation due to uncertainty may vary as well.

Literature most commonly described a test in which the measured performance ratio is compared against a guaranteed performance ratio, with the test lasting 6 months to 2 years. Conformity assessment was framed as a direct comparison against the guaranteed value with an optional inclusion of a guard band at the risk of the final asset owner. A single source instead included the uncertainty through an explicit (2.5%) exceedance probability level of the considered metric. Short-term capacity tests on clear-sky days were also stated to be commonplace.

Performance guarantees for O&M service quality assessment were similarly described as long-term acceptance tests, with a periodic (annual) assessment of the measured performance ratio against a guaranteed value. The guaranteed performance ratio was stated to be decreased over the plant lifetime to account for degradation beyond the O&M provider's control.

The amount of liquidated damages was found to be linked to the net present value of the projected foregone lifetime energy yield due to post-construction underperformance, or the value of the missed energy yield over the testing period in case of O&M underperformance. The mathematical analysis in section 3.3 found the damage clauses for both cases to be separable into a common form: the absolute underperformance of the plant multiplied by a cost factor ξ.

The sector analysis also made apparent some circumstances in which performance guarantees are typically not commonplace, specifically with regards to vertical integration of firm activities along the solar PV value chain. Literature explicitly stated that performance ratio guarantees tend to be less common in the third-party O&M service contractor market, since these contractors are unwilling to assume responsibility for a metric that is in large part determined by construction quality, which they have no control over. Instead, such guarantees in O&M contracts are typically encountered in cases where the EPC and O&M contractors are the same entity or affiliated companies. Similarly, liquidated damages in case of underperformance are only relevant in situations where the asset owner and the service provider are different entities.

The literature study was supplemented by a qualitative research component intended to verify the findings of the literature study, with a secondary objective to potentially obtain new insights. In total, the qualitative research process encompassed 12 informal conversations and 6 formal interviews. Explicit informed consent was only obtained for the formal interviews, therefore results of the informal conversations will not be included in this report. The informal conversations served to inform the research direction during the early stages of this work.

Due to privacy considerations, no identifiable information about the interviewees is made available, nor are transcripts of the interviews publicly accessible. This information was made available exclusively to the university-employed members of the graduation committee.

The firms to which the interviewed professionals belong are active in different sections of the solar PV value chain, listed in table 4-1.

Table 4-1: Company activities of interviewees.

Company 1	Meteorological monitoring system provider, SCADA provider
Company 2	Project developer, EPC, O&M, asset manager/owner, electric utility
Company 3	Project developer, EPC, asset manager/owner
Company 4	Energy yield assessments, SCADA provider
Company 5	Energy yield & performance assessments, uncertainty quantifications
Company 6	Project developer, EPC, asset manager/owner

Of the interviewed companies, 3 are US-based, 2 are located in Europe and 1 is located in India.

In response to being questioned on most common performance assessment practices, the US-based companies mentioned the ASTM E2848 as the most-used performance assessment standard, while the other, non-US, companies stated the PR is primarily used, highlighting regional differences in performance assessment methodologies.

An interviewee from one of the European companies stated that they and other European-based companies employed internal performance assessment procedures instead of strictly adhering to a specific standard, while US-based companies tend to follow the ASTM E2848 standard.

With regards to EPC acceptance testing, this interviewee stated that explicit inclusion of uncertainty of the performance metric, for instance through the form of a P90, is not industry practice. Rather, it was stated that the measured performance ratio and modelled performance ratio are determined under supervision of a third-party engineer, and an acceptance certificate is issued if these two values are within a certain (stated 5%) margin of each other.

This methodology was also outlined by the interviewees of the other European company and the US based companies: the performance requirement is defined as needing to be withing a certain margin of what is calculated by a previously agreed-upon computational model of the power plant. In other words, the performance guarantee is based upon the prediction of a computational model with inclusion of an implicit guard band. The predicted performance ratio is stated to be calculated using the P50 values of the measured environmental data and the modelled yield under these conditions. Multiple interviewees stated that PVsyst is their software of choice for determining performance guarantees.

With the exception of Company 5, that specializes in uncertainty calculations, the measurement uncertainty in the performance metric of choice is determined by an independent engineer that is trusted by all involved parties to the post-construction acceptance test.

One company stated that they possess internal tools for rudimentary uncertainty evaluations to determine P75 or P90 values by applying an uncertainty margin post-measurement, but that this methodology is not used in relation to yield forecasting or acceptance testing.

Although the topic was not discussed with all interviewees, one EPC company that is active in both turnkey as well as build-own-operate PV development projects stated that post-construction acceptance testing is performed in both cases, but that the requirements are significantly less strict in the latter case. This interviewee stated that an analysis of the measurement data is only performed in case the internal performance guarantee is not met. Additionally, this interviewee stated that, in such cases, the investigation may also be directed at potential overperformance in the model and that comparisons may be performed between different yield assessment models when it is suspected that this is the case.

Finally, EPC companies confirmed that liquidated damages in case of underperformance are part of turnkey EPC contracts. One European-based company explicitly confirmed damages being defined as a contractually agreed upon fixed payable amount per $1/10^{th}$ percentage-point underperformance in the PR, thus corroborating what is described in literature. This company also stated that no rewards to the EPC contractor are coupled to overperformance in the PR, a practice that has disappeared due to competition in the EPC industry.

4.2 Uncertainty Calculations

4.2.1 Data Inspection and Cleaning

An initial inspection of the datasets showed that they are both affected by missing data.

- In dataset 1, the data filtering resulted in a selection of 195 out of 365 days suitable for further analysis. Of the 170 days that were deemed unsuited, 160 were missing power output data during the day, 1 contained significant power output during nighttime, 3 contained non-isolated extreme temperature readings, 2 contained significant solar irradiance measurements during nighttime and 4 were excluded upon visual inspection due to uncommonly low overall power output values.
- In dataset 2, the data preprocessing steps resulted in a selection of 297 valid days out of 308 available days. 57 days were excluded from the start due to an inverter malfunction between April 17th and June 12th. Of the 11 days filtered from the data, 4 days were missing output power data during daytime, 5 were missing solar irradiance data during daytime, 4 were missing solar irradiance data and 2 had faulty timestamps leading to incompatibility with the code.

In figure 4.2, the total daily solar irradiation and system power output is shown for each valid day. The gaps in the data can be clearly seen in these figures: dataset 1 has a large gap during November and numerous small gaps throughout the year that seem relatively randomly spaced, whereas dataset 2 has a single large gap from mid-April to mid-June due to the inverter outage, and a few randomly spaced gaps throughout the rest of the year.

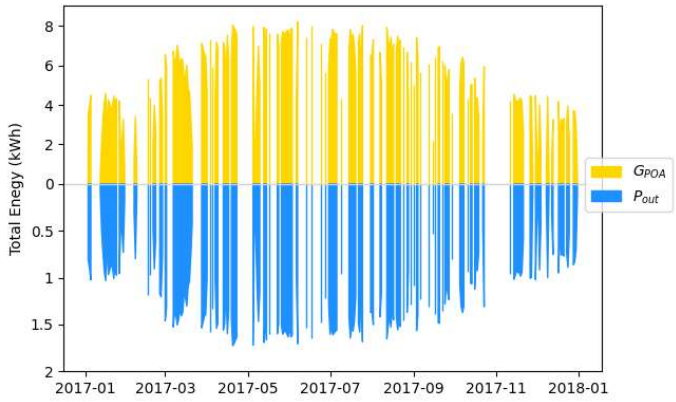


Figure 4.2a: Total daily solar irradiation and power output for valid days of dataset 1.

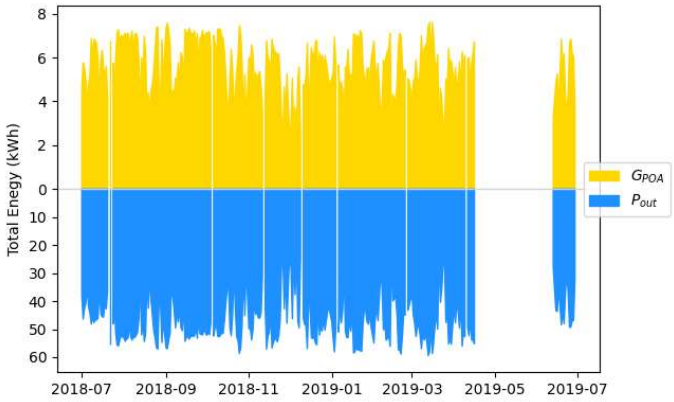
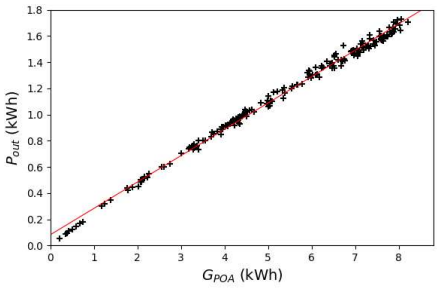
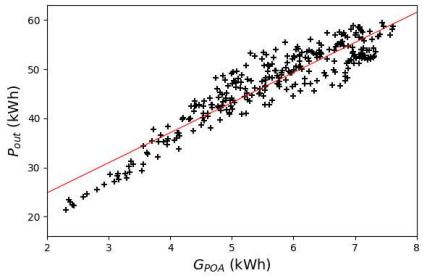


Figure 4.2b: Total daily solar irradiation and power output for valid days of dataset 2.



4.3a



4.3b

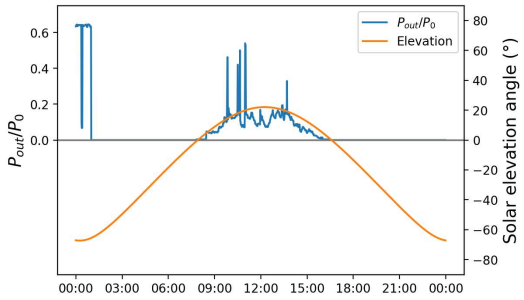
Figure 4.3: Scatterplots of daily total power output against irradiation for dataset 1 (a) and dataset 2 (b). The red lines show the linear least-squares fit to the data.

Visual inspection of figure 4.2 appears to indicate a positive correlation between irradiation and power output, as we would expect. The linear relationship can be seen more explicitly in the scatterplots of figure 4.3.

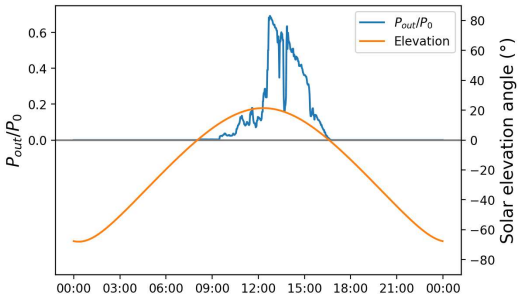
Interestingly, there appears to be a seasonal dependence in the irradiation for dataset 1 and not for dataset 2. A possible explanation for this is the difference in latitude between the two sites; dataset 1 lies at mid-latitude in a continental climate while dataset 2 concerns a system in a tropical climate near the equator.

A selection of graphs illustrating faulty data from dataset 1 is provided in figure 4.4, showing non-zero power output data during the night of 7-12-2017, and zero power output during the morning of 16-12-2017. Both days were excluded from subsequent computations, as are all days caught by the filtering process described in section 3.2.

It might be argued that the early-morning gap in the data as showed in 4.4b does not indicate erroneous data, but rather shading of the module due to an obstacle. However, in such cases one can reasonably expect this shading to occur for multiple consecutive days due to the slow change of the solar diurnal path. Since this is not the case, we cannot confidently dismiss the possibility that the data is incomplete, hence the data is excluded from further analysis.



4.4a: Power output and solar elevation data of 7-12-2017, excluded due to nighttime power generation.

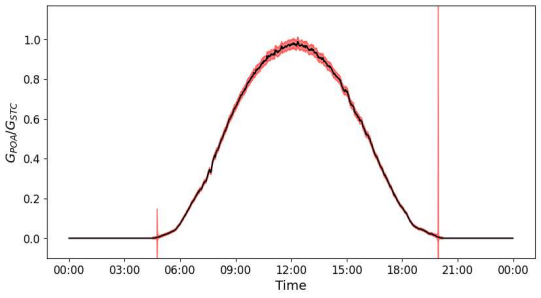


4.4b: Power output and solar elevation data of 16-12-2017, excluded due to zero power generation during the day.

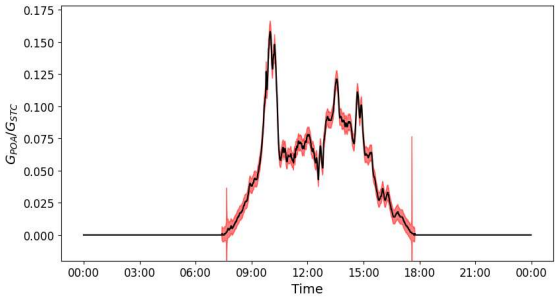
Figure 4.4: Examples of excluded days from dataset 1.

4.2.2 Uncertainty in Performance Ratio Measurements

The 1-minute solar irradiance data including uncertainty (at k=2) for a clear-sky day in late spring and an overcast day in late winter from dataset 1 are presented in figure 4.5. Their associated absolute and relative uncertainties split between the individual pyranometer uncertainty sources are presented in 4.6.



4.5a: Clear-sky day (30-05-2017)



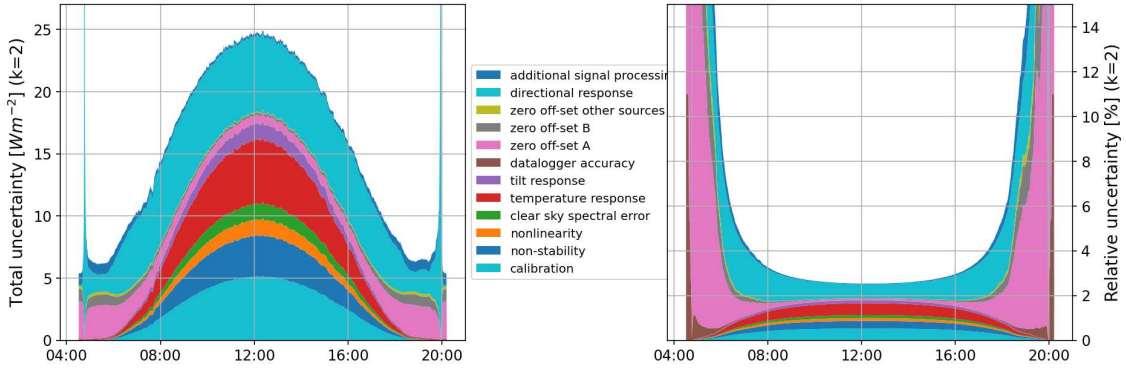
4.5b: Overcast day (08-02-2017)

Figure 4.5: Normalized solar irradiance for two days of dataset 1. Red area indicates uncertainty at level k=2.

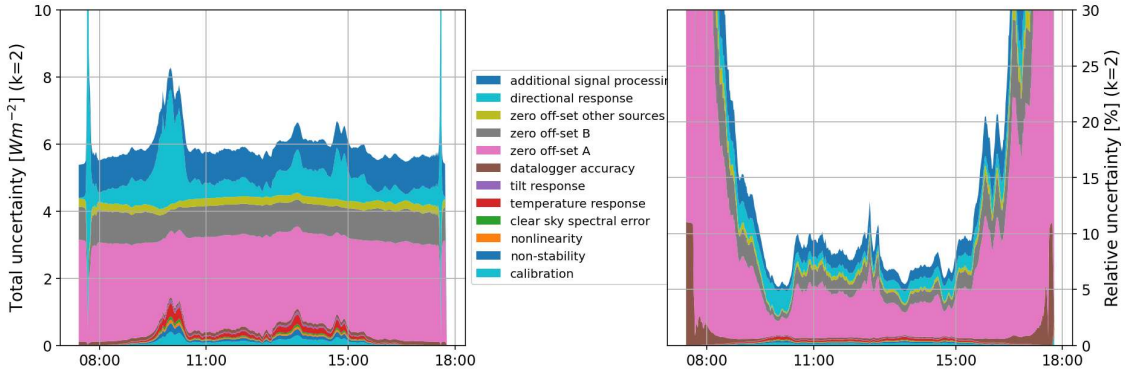
We can observe in figure 4.5 that the maximal irradiance during the overcast day is about a factor 6 lower than the irradiance during the clear-sky day. Total irradiation is 0.6 and 7.8 $kWh \cdot m^{-2}$ for the overcast and clear-sky days respectively, a factor 13 difference. This is in part due to seasonal effects; the total irradiation ratio between an overcast and clear-sky day within the same week of September yields a factor 3 difference, for instance.

The uncertainty splits for the days in figure 4.5 are shown in figure 4.6. As is to be expected, the relative uncertainty sources are significantly less influential to the total measurement uncertainty during days of low irradiation, by virtue of the absolute errors being larger relative to the total signal. During solar noon, the relative uncertainty in irradiance is 2-3 times higher during the overcast day than during the clear-sky day.

The shape of the absolute and relative uncertainty in figure 4.6a roughly corresponds to pyranometer uncertainty calculations found in literature. The absolute uncertainty increases during the day due to the nominal increase of sources with relative bounds, while the relative uncertainty is lowest during times of high solar irradiance and diverges as the irradiance tends to 0 at low solar elevation angles.



4.6a: Clear-sky day (30-05-2017)



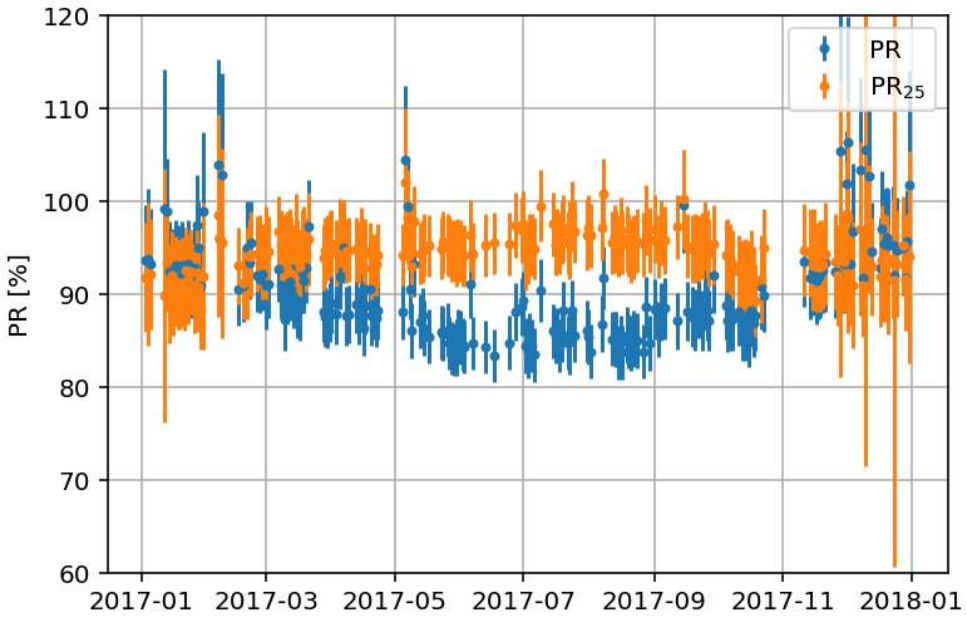
4.6b: Overcast day (08-02-2017)

Figure 4.6: Absolute (left) and relative (right) solar irradiance uncertainty split by root source for a clear-sky day (a) and overcast day (b) from dataset 1.

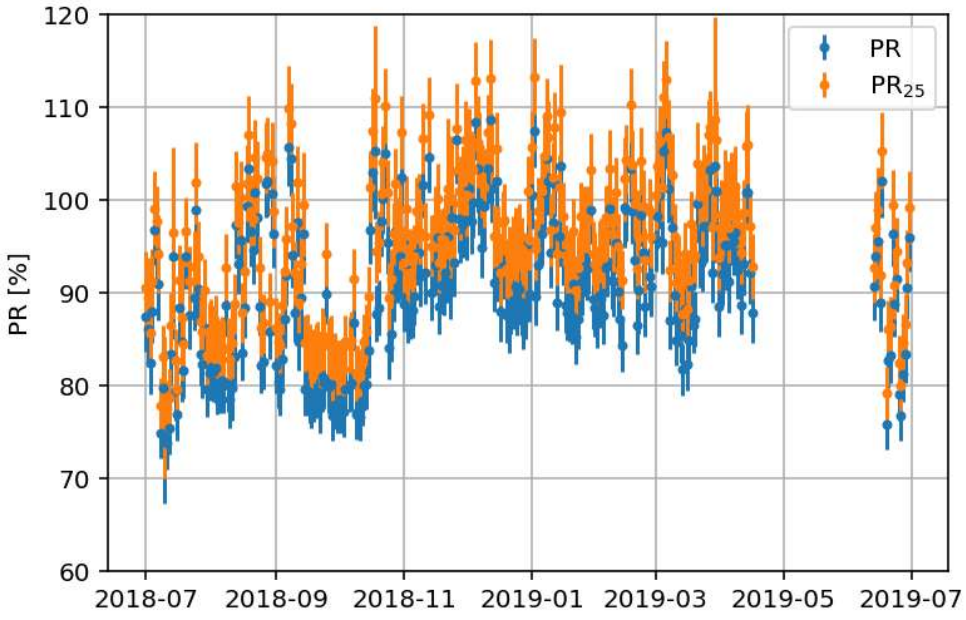
The characteristics in figure 4.6a differ slightly from those in figure 2.5 from Habte et al. (2024) and from the characteristics found by Konings and Habte (2015). The absolute uncertainty in those calculations falls off towards 0 at low solar elevation angles because the directional response error bound is multiplied by a factor $\frac{G_{POA}}{G_{STC}\,cos(\theta)}$, with θ the solar zenith angle (recall from table 2-1 that the directional response error limit is defined relative to a beam radiation with a 1000 $W \cdot m^{-2}$ at normal incidence). The figures in Habte et al. (2024) and Konings and Habte (2015) do not include the divergence of the absolute directional response error as the solar zenith angle tends to 90°, while this divergence can be clearly seen here.

Total daily irradiation uncertainty is 3.23% for the clear-sky day and 10.51% for the overcast day. Total daily irradiation uncertainty is 4.47% for dataset 1 and 3.54% for dataset 2 (k=2).

The daily PR and temperature-corrected PR including uncertainties are calculated for both datasets and presented in figure 4.7. the yearly results are listed in table 4-2.



4.7a: Dataset 1.



4.7b: Dataset 2.

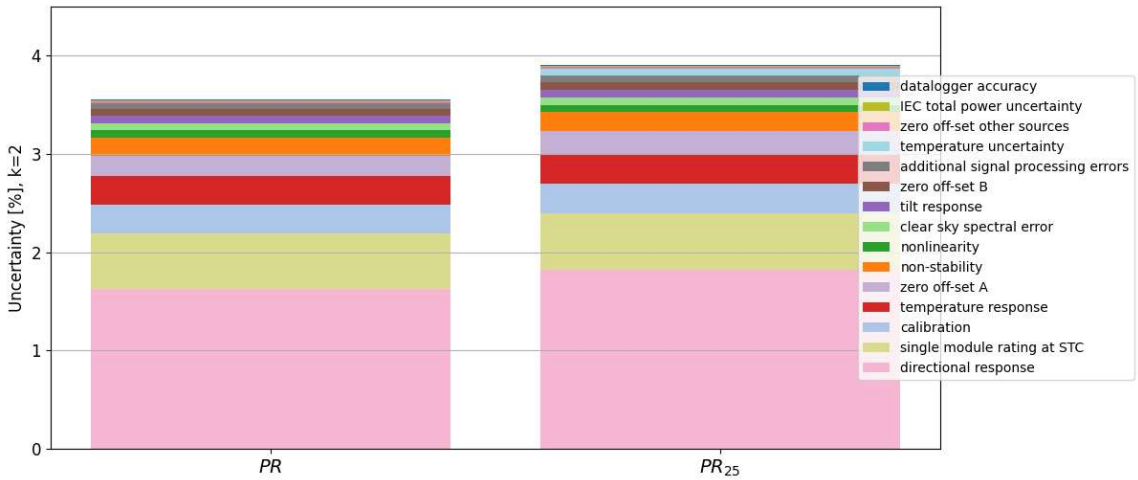
Figure 4.7: Daily PR and temperature-corrected PR values and uncertainties at coverage factor k=2 for both datasets.

For both datasets, the temperature-corrected PR is generally higher than the regular performance ratio. An interesting difference between the datasets is that the temperature corrected PR appears to be structurally higher than the regular PR for dataset 2, while a clear seasonal dependence can be observed for dataset 1. A possible explanation may follow the same logic as the difference in total measured irradiance between the datasets: dataset 1 concerns a system located in a continental climate with distinct temperature differences between summer and winter, while dataset 2 concerns a system in a tropical climate that has high temperatures all-year.

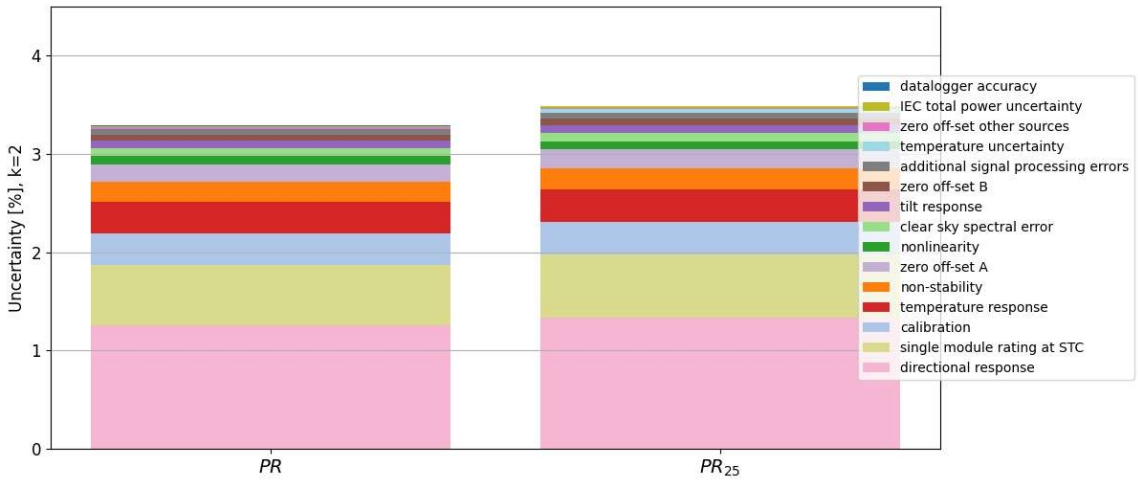
Table 4-2: Yearly performance ratios and expanded uncertainty (k=2) for both datasets.

	PR	PR_{25}
Dataset 1	88.51 ± 3.56 %	94.20 ± 3.91 %
Dataset 2	88.67 ± 3.29 %	93.30 ± 3.49 %

The calculated uncertainties are in line with values reported in literature, which fall between 2-8% at k=2. The contributions to the total uncertainty split between individual sources is shown in figure 4.8. The directional response uncertainty is in all cases the most significant source of uncertainty, being responsible for about 45% of the uncertainty in dataset 1 and for 38% in dataset 2. The module power rating at STC uncertainty, which was defined as normally distributed with $\sigma = 0.8\%$, is the second most significant contributor in both cases, contributing 15% and 18% to total PR uncertainties for datasets 1 and 2, respectively.



4.8a: dataset 1.



4.8b: Dataset 2.

Figure 4.8: Uncertainty in the PR and temperature-corrected PR for datasets 1 and 2, split by root source. Sources are stacked corresponding to their total contribution weight, with the most significant source at the bottom.

In total, around 80% of the uncertainty in the performance ratio estimate is due to pyranometer uncertainty sources. Most significant among these are the directional response (40-45%), temperature response (8-10%), calibration (7-10%), zero-

offset A (7%) and non-stability (5-6%) errors. Uncertainty due to soiling of pyranometer domes was not included in this analysis, however.

These calculations applied the assumption that all uncertainty sources were mutually independent, and fully autocorrelated over time. Complete exclusion of temporal autocorrelation will result in uncertainty estimates of aggregates to tend to 0, due to the fluctuations averaging out over time. Indeed, setting all autocorrelation to 0 results in a (daily) PR uncertainty of about 0.4% at k=2.

In table 4-3 the performance ratio uncertainties obtained using the specification limits for ISO 9060 class B pyranometers, listed in table 3-2, while keeping all other uncertainties constant. Comparing to the results in table 4-2 shows the significant accuracy hit caused by this change. Employing class B pyranometers leads to about 95% of the total PR uncertainty budget being attributable to the pyranometer.

Table 4-3: Expanded uncertainty (k=2) in the (temperature-corrected) performance ratio for both datasets, with pyranometer uncertainty sources set at the ISO 9060 class B specification limits.

	u_{PR}	$u_{PR,25}$
Dataset 1	6.71 %	7.35 %
Dataset 2	6.34 %	6.71 %

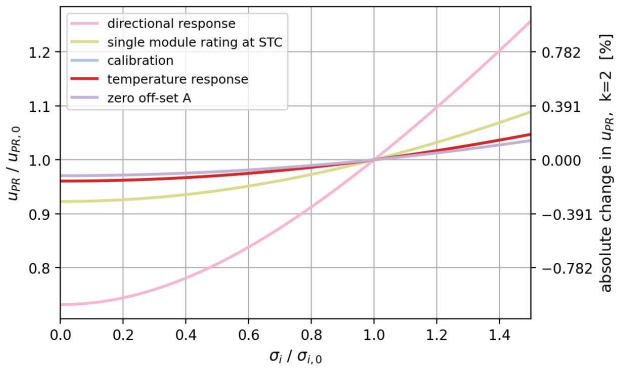
Table 4-4 shows the achieved PR uncertainties for various specification limits. The 'standard' specification limit are the limits outlined in table 3-1, and the uncertainties correspond to those listed in table 4-2. The table further contains results for standard specifications with the directional response limit set to 5 $W \cdot m^{-2}$, results when using the specifications for the Hukx SR300-D1 pyranometer and the results for the SR300-D1 with the directional response limit set to 5 $W \cdot m^{-2}$.

Table 4-4: Expanded uncertainty (k=2) in the (temperature-corrected) performance ratio for both datasets, using different sets of specification limits. Specification 'standard' corresponds to the specifications of table 3-1.

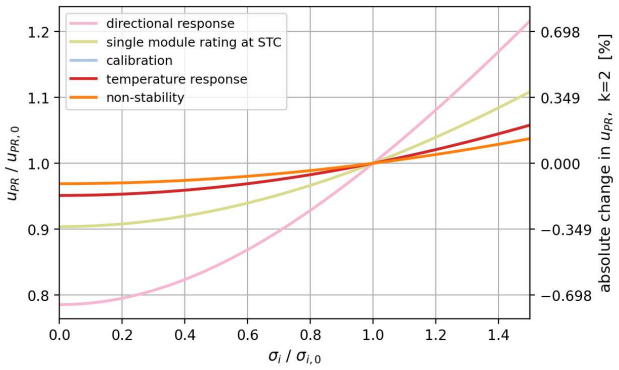
	Standard (values of table 4-2).		Standard with halved directional response		SR300-D1		SD300-D1 with halved directional response	
	u_{PR}	$u_{PR,25}$	u_{PR}	$u_{PR,25}$	u_{PR}	$u_{PR,25}$	u_{PR}	$u_{PR,25}$
Dataset 1	3.56 %	3.91 %	2.88 %	3.15 %	3.06 %	3.39 %	2.24 %	2.48 %
Dataset 2	3.29 %	3.49 %	2.80 %	2.97 %	2.77 %	2.95 %	2.16 %	2.31 %

We can see that, in accordance with the sensitivity analysis shown in figure 4.9, the total PR uncertainty decreases in absolute terms by 0.4 to 0.6% by halving the directional response limit of the ISO 9060 class A specifications. The SR300-D1 is seen to outperform the ISO 9060 class A specifications by 0.5-0.6% as well. The improvement in uncertainty due to a reduction in the directional response error is even more significant for the SR300-D1 pyranometer.

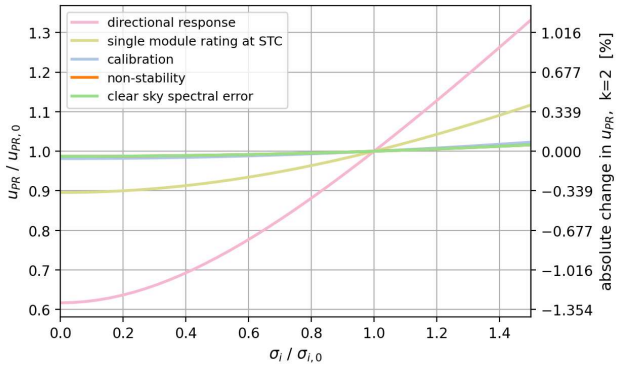
As discussed in the methodology and derived in appendix A.4, the sensitivity of the PR uncertainty to the individual uncertainty sources scales linearly with the relative size of the uncertainty sources when under the assumption that the sources are mutually independent. The explicit relation is given in equation (A.33). The results of the sensitivity analysis of the uncertainty in PR_{25} with respect to the uncertainty specification limit σ_i of the 5 most significantly contributing sources are shown in figure 4.9.



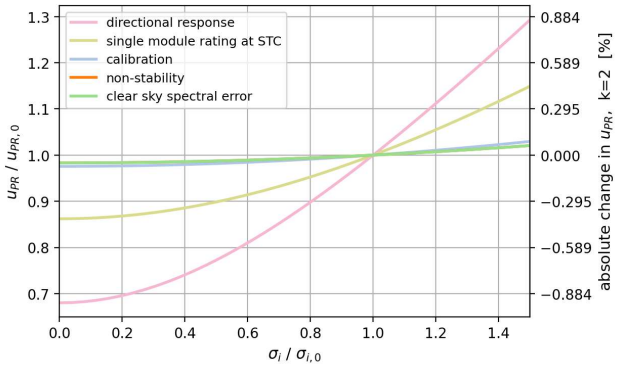
4.9.a1: Standard specifications, dataset 1.



4.9.a2: Standard specifications, dataset 2.



4.9.b1: SR300-D1, dataset 1.



4.9.b2: SR300-D1, dataset 2.

Figure 4.9: Sensitivity of the PR uncertainty to the uncertainty bound σ of the most significantly contributing uncertainties, for the ISO 9060 class A specifications (a) and the SR300-D1 (b). $u_{PR,0}$ and $\sigma_{i,0}$ denote the base values.

The figure further demonstrates that the directional response error is the most significant error source to improve upon in pyranometers. Figures b1 and b2 also show that further avenues for improvement in the SR300-D1 are difficult to identify in the present framework, since the contributions of the other uncertainties are relatively identical. By including more realistic values for autocorrelation and cross-correlation it may become possible to differentiate between these sources and identify further optimal improvement avenues.

4.3 The Impact of PR Uncertainty in Acceptance Testing

The net present value of the foregone yield due to underperformance was determined in section 3.3 to be characterized by a parameter ξ that is equal to the amount of liquidated damages per percentage underperformance. An expression for ξ was derived for the case of post-construction acceptance testing that is dependent on the plant power rating, degradation rate and the average yearly irradiance (in terms of kWh), given by equation 3.11. The value of ξ has been calculated for low, medium and high average irradiation values using the values listed in table 3-4. From the given capex per Watt installed capacity and the total plant power rating, the EPC contract value of the considered power plant is $161 million. The calculation results are presented in table 4-5.

Table 4-5: Liquidated damages owed by the EPC contractor per percentage performance deficiency for different average yearly irradiation scenarios. ξ is expressed in absolute terms and as percentage of total contract value.

	Low irradiation	Medium irradiation	High irradiation
ξ	$601.820	$902.230	$1.203.640
ξ as percentage of the total contract value	0.374 %	0.561 %	0.748 %

The variable ξ sets the scale of the liquidated damages, while the payable amount is determined by the difference between the guaranteed performance and the measured performance. In the above example, we can observe that the damages owed

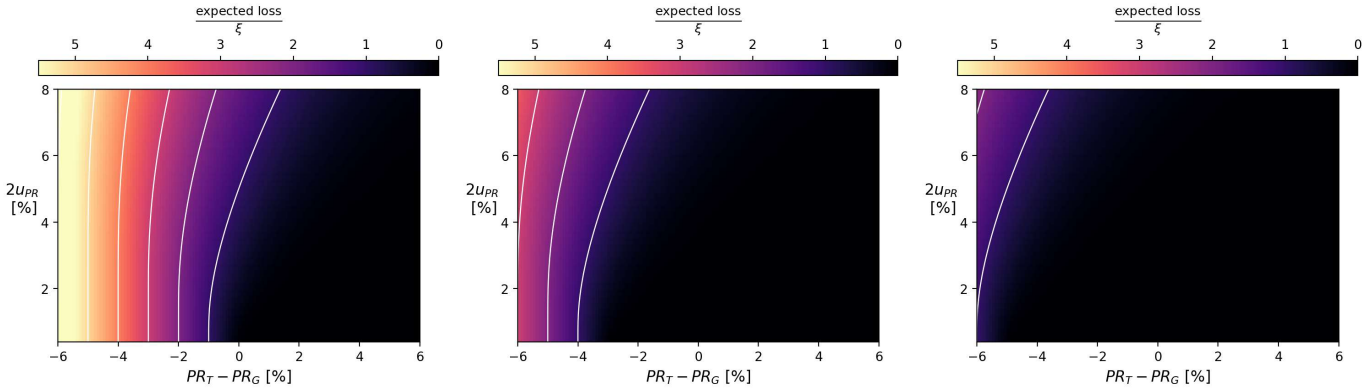
by the EPC contractor to the asset owner will never reach 100% of the contract value, even if the measured PR is 0% and the required PR is 100%. This is because the net present value of the lifetime energy yield is in all considered cases lower than the construction capex; the break-even point for this example lies at approximately 8.02 kWh/m^2/day solar irradiation or a 38 $/MWh levelized PPA price. This is reflected in the fact that the levelized cost of energy presented in the report by Seel et al. (2025) is higher than the levelized PPA price.

The interest of this work is mainly in the second component determining the liquidated damages, the part that is dependent on the measured performance ratio and the specific acceptance testing clause. In section 3.3, the integral equation for the expected damages to which EPC and O&M contractors are liable was derived, which was found to be completely determined by the specific acceptance clause, u_{PR} and the difference between PR_T and PR_G; with EPC and O&M only differing in the definition of ξ.

Specifically, we differentiated acceptance clauses between two cases:

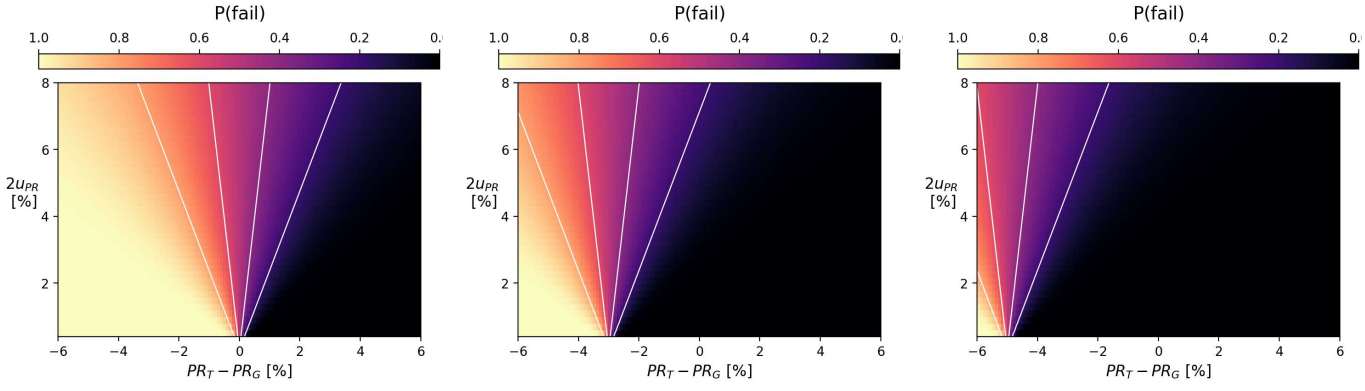
- Case 1: $PR_R = PR_G - \delta$, with $\delta \in \{0\%,\ 3\%,\ 5\%\}$,
- Case 2: $PR_R = PR_G + k \cdot u_{PR}$, with $k = \pm 1, 1.28, 1.96$.

The probability that a power plant does not pass the acceptance criterion, given the unknowable true PR and the measurement uncertainty, is given in equation (3.17) and the expected financial losses incurred by the contractor are given by equation (3.19). These equations have been evaluated for a domain of measurement uncertainties and possible values for PR_T. The results are presented for each case in figures 4.10 to 4.12. Uncertainties are presented at k=2.



4.10.a1: $\delta = 0$ *4.10.a2*: $\delta = 3\%$ *4.10.a3*: $\delta = 5\%$

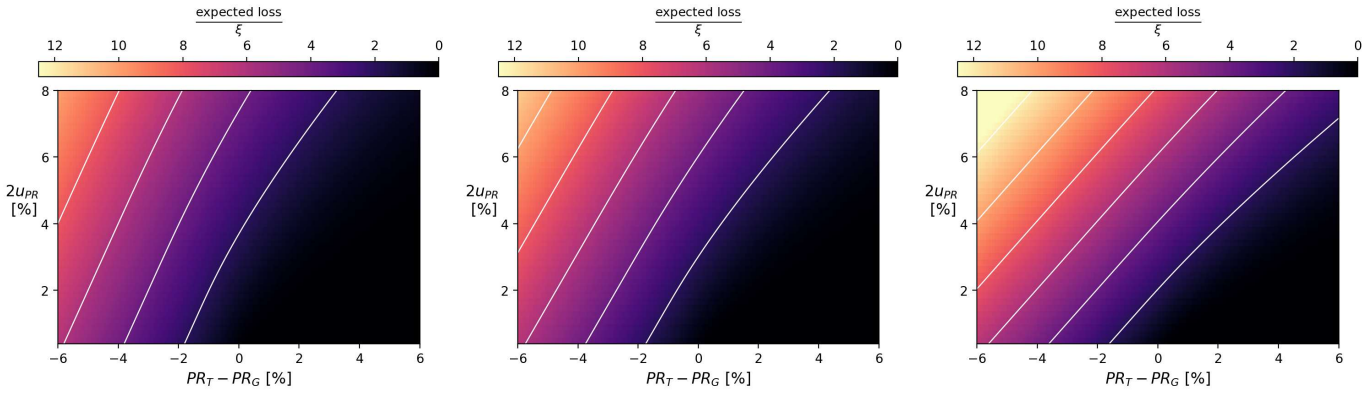
4.10a: Expected loss in units of ξ for acceptance testing of type 1.
White lines trace the contours of magnitude 1, 2, 3, 4 and 5.



4.10.b1: $\delta = 0$ *4.10.b2*: $\delta = 3\%$ *4.10.b3*: $\delta = 5\%$

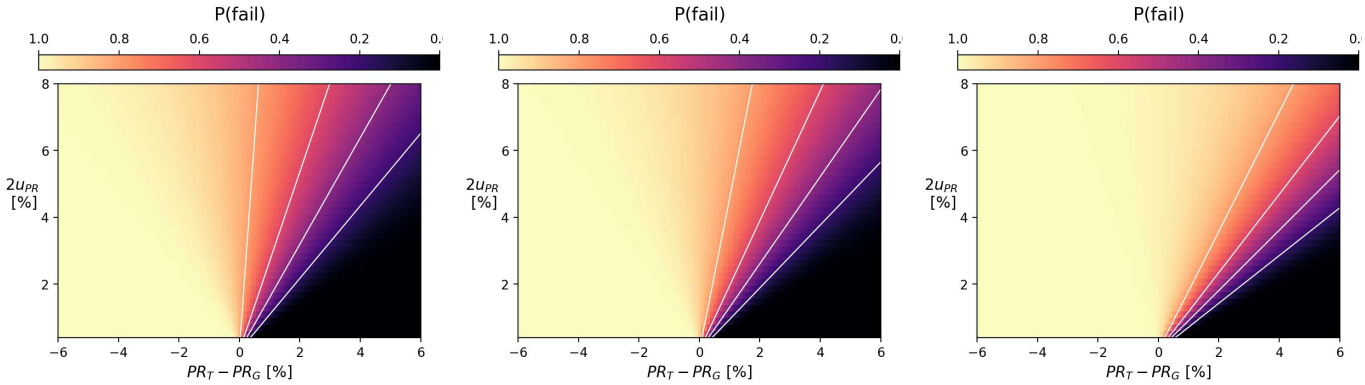
4.10b: Probability of to meet the required PR in acceptance testing of type 1.
White lines trace the 0.2, 0.4, 0.6 and 0.8 probability contours.

Figure 4.10: Expected payable damages (a) and probability to of failure to meet the required PR (b) for acceptance testing clause of type 1 for three different guard band parameters δ. The guard band causes a translation of the characteristic to the left, shifting the risk to the asset owner.



4.11.a1: $k = 1$ *4.11.a2:* $k = 1.28$ *4.11.a3:* $k = 1.96$

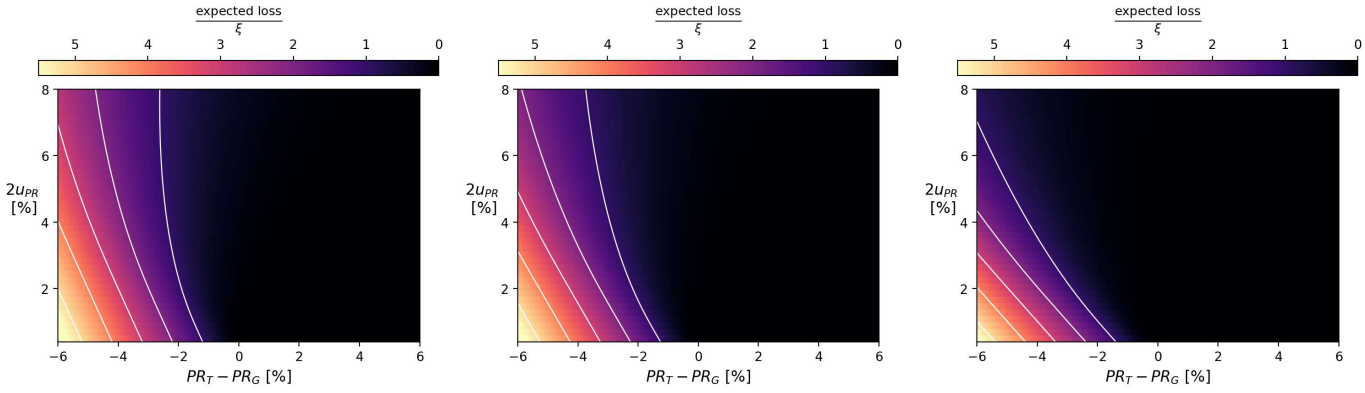
4.11.a: Expected loss in units of ξ for acceptance testing of type 2 with positive k.
White lines trace the contours of magnitude 2, 4, 6, 8, 10 and 12, which are much higher than those of figures 4.10a and 4.12a.



4.11.b1: $k = 1$ *4.11.b2:* $k = 1.28$ *4.11.b3:* $k = 1.96$

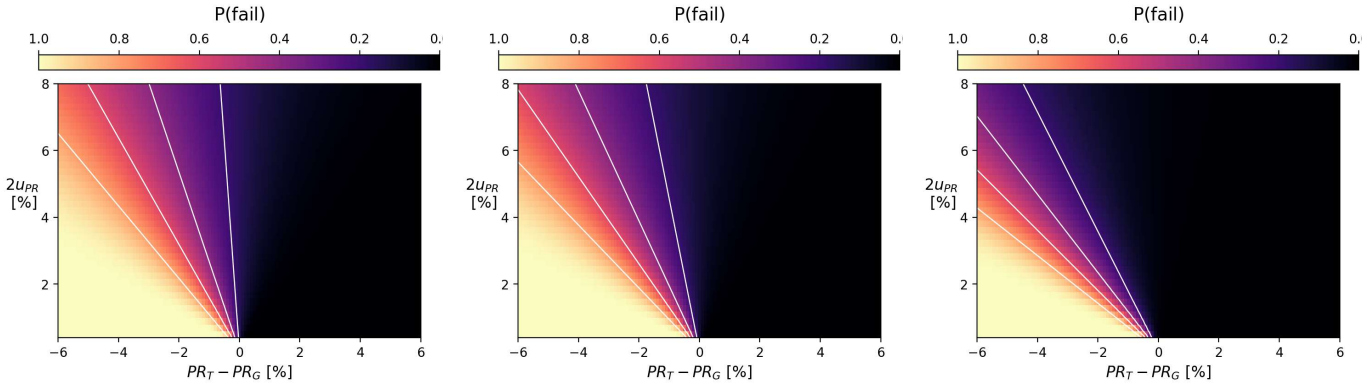
4.11b: Probability of to meet the required PR in acceptance testing of type 2 with positive k.
White lines trace the 0.2, 0.4, 0.6 and 0.8 probability contours.

Figure 4.11: Expected payable damages (a) and probability to of failure to meet the required PR (b) for acceptance testing clause of type 2 for three different values of positive k, fixing PR_R at the P84, P90 and P97.5 exceedance levels for PR_G. With increasing k, risk is shifted to the EPC or O&M contractor (since the exceedance level is increased). Observe the higher expected damages compared to the other figures.



4.12.a1: $k = -1$ *4.12.a2:* $k = -1.28$ *4.12.a3:* $k = -1.96$

4.12a: Expected loss in units of ξ for acceptance testing of type 2 with negative k.
White lines trace the contours of magnitude 1, 2, 3, 4 and 5.



4.12.b1: $k = -1$ *4.12.b2:* $k = -1.28$ *4.12.b3:* $k = -1.96$

4.12b: Probability of to meet the required PR in acceptance testing of type 2 with negative k.
White lines trace the 0.2, 0.4, 0.6 and 0.8 probability contours.

Figure 4.12: Expected payable damages (a) and probability to of failure to meet the required PR (b) for acceptance testing clause of type 2 for three different values of negative k, fixing PR_R at the P16, P10 and P2.5 exceedance levels for PR_G. With decreasing k, risk is shifted to the asset owner (since the exceedance level is decreased).

Recall that literary sources list PR uncertainty between 2-8% (k=2) and the calculations performed in this work yielded expanded uncertainties between 3.3-3.9% when using standard ISO 9060 class A specifications, and 2.8-3.4% when using SR300-D1 specifications, as listed in table 3-1.

From a statistical point of view, the expected payable damages ideally follow perfect vertical contours, in which case the expected liquidated damages correspond exactly to the true loss due to underperformance. In other words, vertical contours of expected damages correspond to a situation where there is no statistical bias in favor to either the asset owner or the contractor. Similarly, the probability that the EPC or O&M contractor owes damages should be 1 if $PR_T < PR_G$ and 0 if $PR_T > PR_G$. Since measurement uncertainty causes a broadening of the distribution this is mathematically impossible by definition.

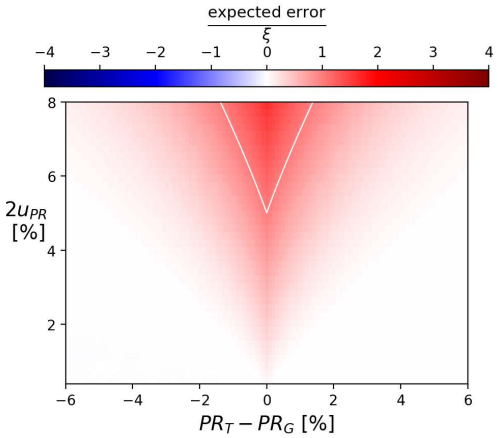
As is to be expected, acceptance criteria of type 2, where acceptance is coupled to a specific exceedance probability, shifts the risk significantly to one party. The expected liquidated damages are especially skewed for positive k, with the contractor being liable for approximately 2ξ over the fair amount in the uncertainty range of 3-5%. Additionally, failure to meet the acceptance criterion is significantly nonzero even in cases where the PR_T significantly exceeds the guarantee, due to PR_R shifting upward with the uncertainty magnitude. We can see from figure 4.11b that the probability of making a type-2 error (false rejection) increases with the uncertainty.

The reciprocal situation is shown in figure 4.12. While the rejection probability is perfectly inverted (leading to an increased chance of type-1 errors), we see that the characteristics of the expected payment are far more vertical in this case. The difference between the two situations is that positive k results in expected payments in the regime where no payments should be due at all, while negative k leads to lower, but nonzero, expected payments in a regime where payments are due. Thus, the overall payment error is lower in this case.

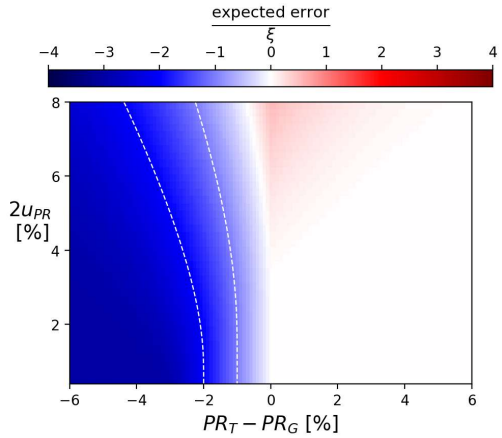
Meanwhile, acceptance criteria of type 1 are far more symmetric in the failure probability distribution. An increase of δ simply constitutes a translation of the entire characteristic to the left (i.e. shifts the risk to the asset owner). We can observe that the case with δ=0 is overall the most fair type 1 acceptance criterion in the specified parameter range.

While acceptance type 1 with δ=0 results in a perfectly balanced probability of false acceptance and false rejection (see figure 4.10.b1), the expected damages are not necessarily fair, since the risk due to higher measurement uncertainty is exclusively at the cost of the EPC or O&M contractor. As a measure to investigate fairness, figure 4.13 presents the error between the expected liquidated damages and the true damages for four different cases. It can be readily observed that

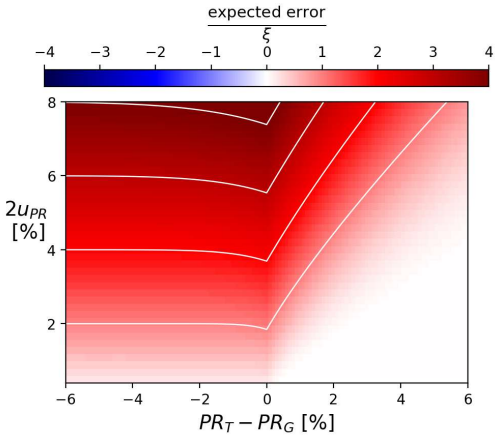
acceptance testing of type 1 (with $\delta = 0$) leads to the least overall payment error, but the expected error is entirely at the cost of the contractor. Inclusion of non-zero guard bands constitutes a transfer of the risk to the asset owner and a sharp increase in the error. While simple acceptance test of type 1 with $\delta = 0\%$ appears the fairest in this case, it should be acknowledged that the specific contractual agreement is a topic of extensive negotiation in which either party wants to shift the risk to the other. The mathematically 'fairest' agreement is therefore not necessarily the most common outcome.



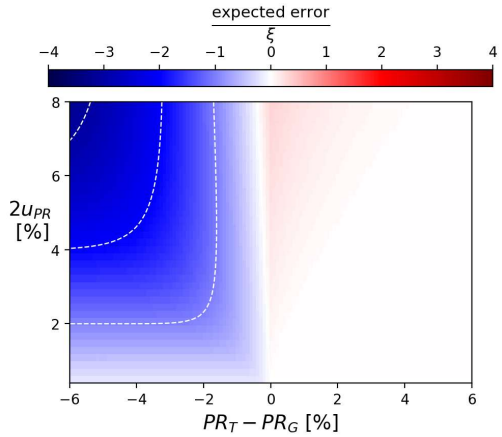
4.13a: Acceptance type 1, $\delta = 0\%$.



4.13*b:* Acceptance type 1, $\delta = 3\%$.



4.13c: Acceptance type 2, $k = 1$.



4.13d: Acceptance type 2, $k = -1$.

Figure 4.13: Error between expected damages and true damages for various acceptance criteria. Red indicates the expected damages are too high, blue indicates they are too low. White (dashed) lines trace the contours of positive (negative) integer-valued error.

From the results in figures 4.10 to 4.13 we can observe that the implications of measurement uncertainty differ between acceptance testing types. In all cases, the asset owner statistically benefits from measurements with lower accuracy due to the nonlinearity of the payment function. This is clearly seen in figure 4.13a: the EPC or O&M contractor statistically overcompensates the asset owner, with the overcompensation increasing with the uncertainty. While omnipresent, the effect is superposed with the more substantial effects of guard banding or exceedance probability acceptance testing.

Inclusion of a guard band significantly shifts the financial risk to the asset owner, as seen in figure 4.13b. However, the risk-shift at the benefit of the asset owner remains clearly visible as the uncertainty increases.

In tests of type 2 with positive k, the statistical advantage of the asset owner is amplified by the fact that higher uncertainty increases PR_R. Because of these stacking effects, type 2 tests with positive k are statistically the most unfair type of acceptance test, with significant dependence on the measurement uncertainty at the advantage of the asset owner.

Conversely, higher uncertainty decreases PR_R for negative k, which counteracts the asset owner's inherent advantage. We can see in figure 4.13d that the expected payable damages are in the favor of the EPC or O&M contractor in this case. This is

the only case where higher uncertainty is advantageous for the EPC and O&M contractors, and even then exclusively in the regime where the degree of underperformance is greater than the measurement uncertainty. As the uncertainty increases, PR_T and PR_G become approximately equal on the scale of the uncertainty and the probability distribution $\mathrm{P}(PR_M - PR_R)$ converges, leading to statistical independence of the expected damages to the measurement uncertainty. Thus, the benefit for the EPC and O&M contractors is less than the benefit of the asset owners in the case of positive k.

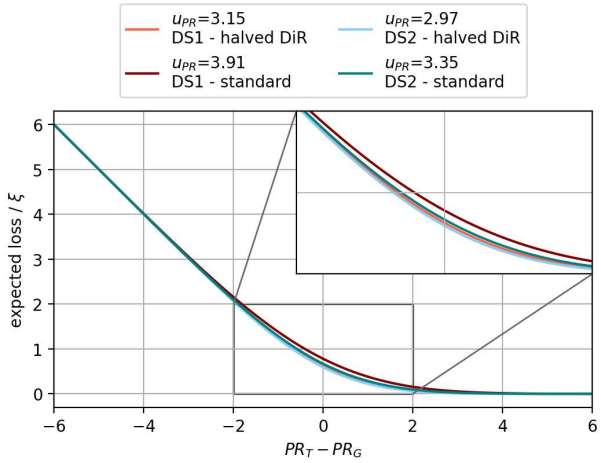
From the interviews it was gathered that acceptance tests of type 1, potentially with guard bands, are most commonly encountered in industry. While measurement uncertainty works in favor of the asset owner, guard banding practices shift the balance more significantly to the favor of the EPC contractor. Thus, uncertainty reductions statistically benefit EPC and O&M contractors, but a potential commensurate reduction of the guard band will be in the interest of the asset owner.

While the effect of uncertainty reductions on the practice of guard banding is difficult to predict, we can quantify the effect of lower measurement uncertainty in the existing financial risk of EPC and O&M contractors. In figure 4.14, the effect of lower measurement uncertainty to the financial risk assumed by the EPC and O&M is shown for different uncertainty reductions in type 1 guarantees with $\delta = 0\%$. The figure shows the horizontal cross-sections of figure 4.10.a1 at the PR uncertainty levels in table 4-4. The figures are comparisons between different uncertainty specifications for pyranometers:

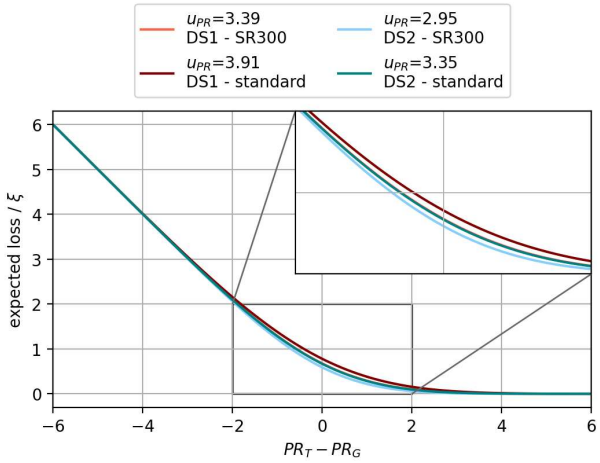
a) Standard uncertainties (table 3-1, ISO 9060 class A pyranometers) compared to standard specifications with halved directional response uncertainty.
b) Standard uncertainties compared to the Hukx SR300-D1 pyranometer specifications.
c) Standard uncertainties compared to the SR300-D1 with halved directional response uncertainty.
d) The SR300-D1 specifications compared to the SR300-D1 with halved directional response uncertainty.

The data shows that the measurement uncertainty influences the expected payable damages only when PR_T is close to PR_R. In the situation where PR_T is much higher or lower than the requirement the expected damages are not sensitive to the measurement uncertainty and follow the ideal (flat and diagonal, respectively) characteristics. This is mathematically explained by the probability distribution being almost completely inside the acceptance or rejection domains for PR_M at the given uncertainty level. Note that the width of the regime where the expected damages deviate from the ideal broadens as the uncertainty increases, as clearly shown in figure 4.10.a1

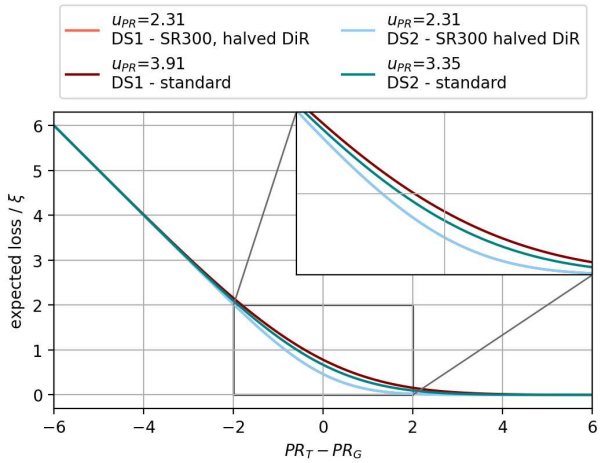
The insets in the figures provide more detailed depictions of the regime where measurement uncertainty has the greatest influence. We can nicely see the impact of a directional response error reduction in figure 4.14a, with the expected damages decreasing by about 10 to 20%, or by about 0.15ξ in absolute terms. The SR300-D1 provides a similar (12%) improvement over the standard ISO 9060 class A specifications. The SR300-D1 with a halved directional response error achieves a reduction of 30-35% in the expected payable damages.



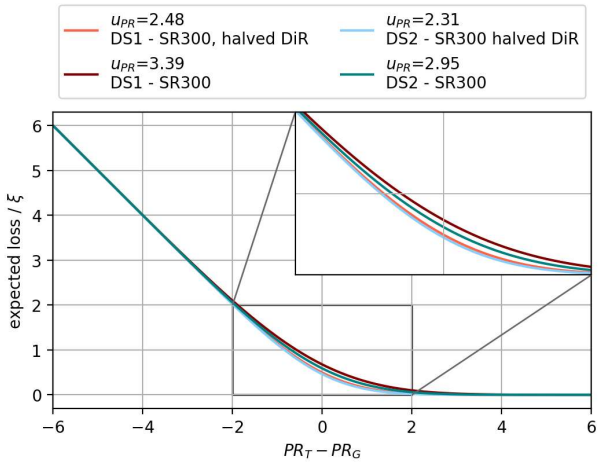
4.14a: Comparison of standard specifications to those with halved directional response specification.



4.14b: Comparison of standard specifications to the specifications of the SR300-D1.



4.14c: Comparison of standard specifications to the specifications of the SR300-D1 with halved directional response.

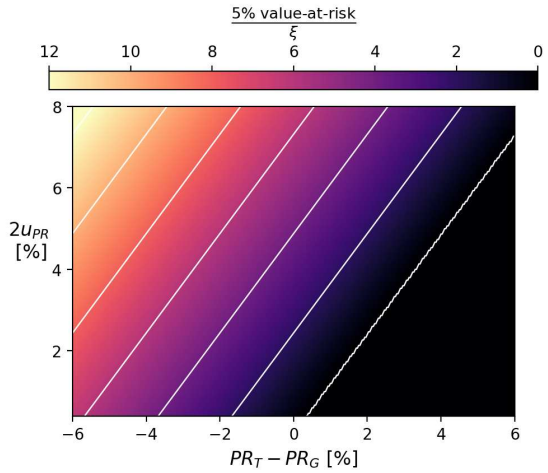


4.14d: Comparison of SR300-D1 to the SR300-D1 with halved directional response.

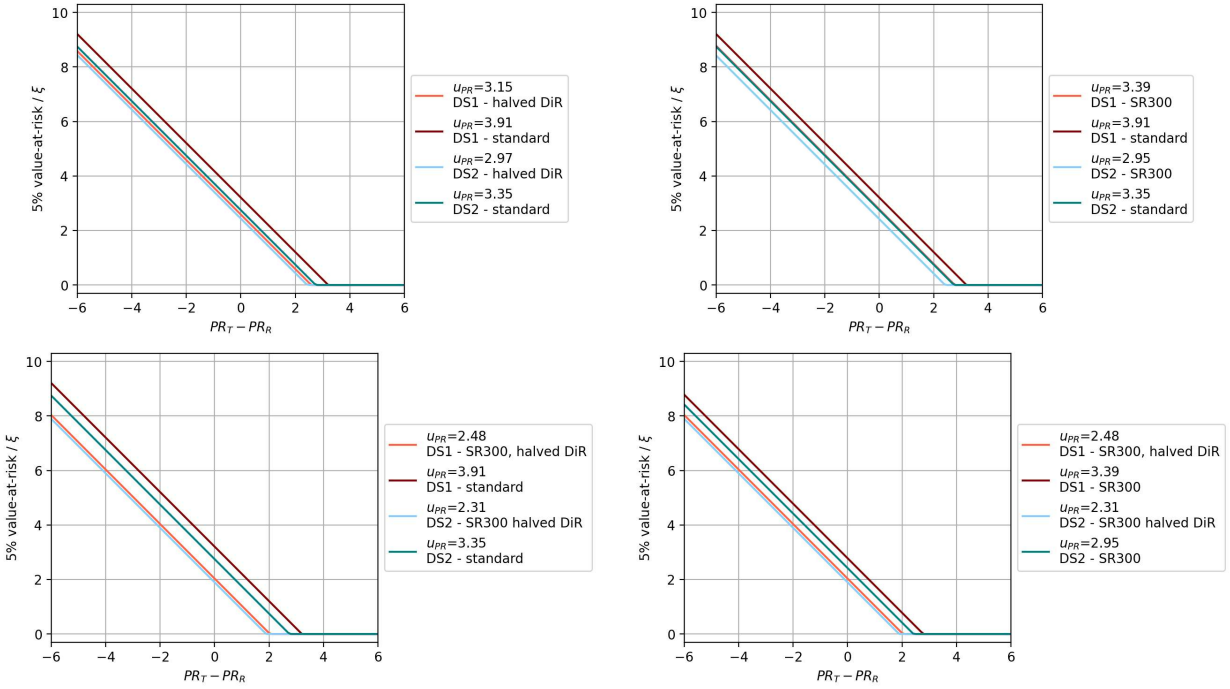
Figure 4.14: Expected liquidated damages against the true plant performance relative to the guaranteed value, for different values of u_{PR} (k=2) given by table 4-4. Figures concern performance acceptance testing of type 1 with $\delta = 0\%$. The red lines concern results for dataset 1 (DS1), the blue lines results for dataset 2 (DS2).

Clearly, the most effective way to reduce the expected payable amount of liquidated damages is by ensuring the PR_T is significantly greater than the required value, by a margin of u_{PR}. However, it is important to keep in mind that PR_T is an instantaneous performance ratio that would be obtained using perfect measurement instruments, and that there are numerous factors such as degradation, soiling or partial shading of the power plant that affect its value. Furthermore, there are physical limitations to the performance that can be achieved through better construction or maintenance work. Hence, a buffer may be difficult to achieve, or at least ensure, in practice. It may therefore be worthwhile to invest in higher accuracy measurement equipment to reduce the potential liquidated damages. Recall that ξ was calculated to exceed $1.2 million for a utility-scale plant in a location with high average irradiation, meaning a reduction of 0.15 ξ equals a reduction of expected damages of $180.000.

Furthermore, we have been discussing the *expected amount of payable damages*, which concerns a statistical expectation value. In financial risk assessment, another metric of interest is the *value-at-risk* defined in equation (3.20) which provides insight in the distribution's tail-end of worst-case outcomes. The value-at-risk is the upper bound of the loss incurred in the worst $\alpha - \%$ of cases. In figure 4.15, the heatmap and cross-sections of the 5% value-at-risk in acceptance testing of type 1 with $\delta = 0\%$ are shown.



4.15a: Heatmap of the EPC or O&M contractor value-at-risk at the 5% level. Linear behavior of the characteristic can be readily observed.



4.15b: Cross sections of the value-at-risk graph for various uncertainty improvements.

Figure 4.15: EPC or O&M contractor value-at-risk at the 5% level for acceptance testing of type 1 with $\delta = 0\%$.

While the expected liquidated damages are only dependent on the uncertainty in a band around the performance requirement and statistically regress towards the mean at very high and low construction quality, we see that the value-at-risk does not exhibit such behavior due to it being tied to the measurement uncertainty by definition. From figure 4.15b, we see that halving the directional response specification can reduce the value-at-risk for the contractor by 0.3ξ to 0.75ξ. Hence, even in situations of significant underperformance, the financial risk is reduced by more precise measurement.

5. Discussion

5.1 Results and Methodology

The literature study performed during this research revealed that published studies investigating uncertainty in PV plant performance metrics include pyranometer uncertainty as a flat percentage, while this uncertainty is in reality composed of multiple individual uncertainty sources with their own characteristics. This work adds to the existing literature by performing uncertainty calculations of the performance while keeping these contributions explicit, with the purpose of identifying the most impactful avenues for pyranometer improvement. These results were furthermore placed in the context of the practical applications of PV plant performance metrics.

PV system performance metrics are used for multiple purposes during a PV power plant's operational lifetime, for instance as a generic performance indicator or tool to determine asset value. A notable application of performance metrics is as an assessment method to determine quality of construction and maintenance work performed by external EPC and O&M contractors. Contractual agreements for EPC and O&M services contain clauses specifying a guaranteed value of a performance metric that must be achieved during a test conducted at an agreed-upon moment. In case the guaranteed level is not reached, the contractor is required to pay liquidated damages to the asset owner as reimbursement. These damages were calculated to be proportional to the measured underperformance. This procedure was confirmed as common practice by industry professionals.

Thus, a direct connection was established between measurement uncertainty in performance metrics and the financial interests in contractual performance guarantees, through which pyranometer accuracy improvements could be contextualized.

The literature study found that there exist two authoritative standards outlining performance monitoring and assessment methods for solar PV power plants, the IEC 61724 and the ASTM E2848 standards. These standards define multiple different performance metrics, but this work restricts itself to the performance ratio, or PR, defined in the IEC 61724 standard. The PR was found to be commonly applied in the European PV industry and has the notable feature that it only depends on measured data and not on computational models, which makes it highly suited for uncertainty propagation calculations. Previous works calculating uncertainty in performance ratio achieved expanded uncertainties between 2-8%.

The uncertainty in the regular and temperature-corrected performance ratio was calculated for two datasets, obtaining expanded uncertainties of 2.8% to 3.9%, in line with published estimates. These estimates were obtained by propagating uncertainties according to the GUM methodology. Pyranometer uncertainties were taken as the ISO 9060 class A specification limits and as the specification limits of a state-of-the-art industrial class A pyranometer (the Hukx SR300-D1). Uncertainties in the other variables were taken as standardized uncertainty limits where specified by the IEC 61724 standard, and taken from literature otherwise.

The SR300-D1 was found to significantly outperform the standard class A specifications, providing a reduction of 0.6% in the PR uncertainty, a relative accuracy increase of over 15%. This demonstrates that a range in quality can exist within current standardized classes, since the SR300-D1 is officially categorized as a class A pyranometer. Thus, if accuracy is of concern in a certain pyranometer application, it is insufficient to rely exclusively on an ISO 9060 categorization given the improvement that state-of-the-art instruments can provide over instruments following class requirements.

This work assumed all uncertainty sources to be independent and fully autocorrelated over time. The independency assumption allowed for an analytical decomposition of the total uncertainty to the contributions of individual uncertainty sources. Furthermore, it provided a simple expression for the sensitivity of the total uncertainty to the uncertainty of each

constituent source. Calculations showed that 75-85% of the total uncertainty in the PR was attributable to the pyranometer measurement and calibration uncertainty, with the PV module calibration being the second most significant source. Power metering uncertainty and temperature sensor uncertainty were found to have a marginal effect on the total uncertainty.

The directional response error was found to be the most significant contributor in all cases, being responsible for 35-60% of the total uncertainty budget. The uncertainty due to the module calibration contributed 15-30%. The class A specifications also resulted in significant contributions of the temperature response (8-10%), pyranometer calibration (7-10%) and zero offset A (6%).

The SR300-D1 significantly outperforms the ISO 9060 class A specifications for these latter three sources, reducing their contributions to comparable significance as other pyranometer uncertainties, leaving the directional response error as the single most important source of uncertainty in the SR300-D1. Efforts for pyranometer improvement should therefore be focused on reducing the directional response error.

Important to note is that the directional response uncertainty is specified with respect to the incoming direct irradiance. The data used for the calculations lacked direct irradiance measurements and thus the bound was calculated using the global irradiance, leading to an overestimation of the directional response error by 15% or higher, depending on cloud conditions. Therefore, an effective method to reduce measurement uncertainty using currently available technology is to combine the pyranometer measurements with pyrheliometer or diffusometer data to obtain a correct error bound for the directional response. Note that this only provides an ex-post reduction of the measurement uncertainty bound (i.e. an increase of the P90 irradiance) and does not affect the measured irradiance value (the P50 irradiance).

The assumption of independent uncertainty sources constitutes a potential underestimation of the total uncertainty due to neglection of cross-covariance contributions. Conversely, assuming complete autocorrelation results in an overestimation of the uncertainty due to noisy sources by treating them as an offset. Without proper characterization of these covariances, no indication can be given on the magnitude of these errors. It might be argued that an over-estimation of the uncertainty would generally be preferrable, which would be reflected through full cross-correlation between sources inside the same physical component. Such a treatment would combine all pyranometer uncertainty sources into a single offset, which has already been the approach of previous studies. The present methodology helped identify the directional response as a significant error source. A covariance characterization would be required to identify further optimal improvement avenues, since the contributions and sensitivities of remaining pyranometer uncertainty sources are roughly identical for the SR300-D1 under the current assumptions.

The impact of uncertainty in PR was studied in the context of performance guarantee conformity assessment in EPC and O&M contracts. Two types of performance guarantee clauses were identified from literature. Type 1 clauses simply require the measured PR to be higher than the guaranteed PR, or the guaranteed PR minus a guard band (typically 3-5%). Type 2 clauses require the measured PR to exceed the guaranteed PR at a certain exceedance probability level. Clauses of type 1 were the more commonly encountered type in literature, and interviewed professionals from industry confirmed that type 1 clauses are common practice. The interviewees also stated that the guaranteed PR is typically defined as the PR calculated by a computational model of the power plant.

The influence of uncertainty to performance guarantee conformity assessment was quantified by calculating the acceptance probability and the expectation value of the liquidated damages for a range of situations. The calculations rely upon a ‘true PR’ (the PR that would be measured by perfect measurement instruments) that is measured with a normally distributed measurement uncertainty.

As a consequence of the nonlinearity of the liquidated damages (equation 3.12), measurement uncertainty was seen to statistically increase the expected payable damages in almost all cases, increasing the financial risk for the EPC and O&M contractors. The effect was seen to be small in comparison to the risk-shifts arising from guard bands and exceedance-

probability-based acceptance tests. Hence, it is overshadowed in type 2 acceptance tests with $k < 0$, leading to higher measurement uncertainty becoming beneficial to EPC and O&M contractors.

An assessment of the difference between the expected liquidated damages and the fair liquidated damages (i.e. the expected error in owed damages) showed that implementing guard bands or coupling acceptance to exceedance probability levels resulted in significant errors, i.e. unfair risk-shifts to one involved party. The lowest overall error was, perhaps unsurprisingly, achieved when the requirement is a simple exceedance of the guaranteed PR by the measured PR. However, the expected error was found to be fully at the expense of the EPC or O&M contractor in this case. Guard bands were seen to significantly shift financial risk over to the asset owner. Overall, lower uncertainty combined with smaller guard bands leads to statistical fairness of acceptance tests.

For type 1 tests, it was observed that the influence of measurement uncertainty to the expected liquidated damages is only significant in cases where the true PR is close to the contractual requirement. The width of this domain is characterized by the magnitude of the uncertainty itself. However, an isolated performance assessment is still affected by measurement uncertainty, which may cause the damages to be far higher or lower than the true value; the error only averages out statistically. This was reflected by the linear relation between the uncertainty and the value-at-risk. By halving the directional response error, the 5% value-at-risk could be reduced by up to 0.75ξ. In a post-construction acceptance test of a theoretical 100MW power plant this was shown to correspond to $900.000.

In the aforementioned regime where the true PR is close to the requirement, pyranometer uncertainty has a significant impact on the expected liquidated damages. In this regime, the difference between standard ISO 9060 class A specifications and the SR300-D1 yields a decrease in the expected damages by 13%. If the directional response error of the SR300-D1 were to be halved, this reduction would approach 35%, or 0.3ξ. For the 100MW power plant, the nominal value of this reduction would be in the order of $360.000.

Calculations showed that the most effective way for an EPC or O&M contractor to reduce financial risk (aside from negotiating large guard-bands before signing a contract) is by ensuring the true PR is significantly higher than the requirement. However, this true PR remains an instantaneous quantity that is dependent on unpredictable factors such as the weather, degradation rates and soiling. Increasing performance through superior construction or maintenance quality eventually reaches physical limitations and hence may not be a feasible approach. It was shown that high-accuracy pyranometers such as the SR300-D1 can provide a substantial reduction to the financial risk in acceptance testing in terms of the expected liquidated damages and the value-at-risk, even compared to the present class A standard. The risk can be reduced even further by improving the directional response of pyranometers.

Example calculations in which the acceptance criterion is optimized to reduce total error are provided in appendix B. There are two reasons why these calculations are not included in the main results section. First of all, no deliberate methodology was designed or followed in these calculations, while the optimized performance metric are highly sensitive to the chosen domain over which the error is optimized. Confident recommendations on optimal metrics would require more substantiative justification of the assumptions. Secondly, and more significantly, the interviews and sector analysis cast doubts on whether industry is prepared to change common legal practices to base performance guarantees and liquidated damages on untransparent mathematically optimized criteria. Especially criteria that are dependent on ex-post uncertainty estimates may become points of contention between the involved parties.

Even if a criterion could be constructed that minimized overall error in liquidated damages or equally distributed financial risk between both involved parties at a high confidence, it may not be adopted by industry. EPC and O&M contractual specifications are the result of historical experience and negotiations between involved parties. Statistical optimality of risk allocation may be far outweighed by asymmetrical power dynamics between the negotiating parties. The optimization of the performance metric should therefore be regarded as a proof-of-concept with limited practical applicability.

The calculations of the statistical error in the acceptance testing clauses found in literature are aimed to provide insight in the fairness of different possible practices. The results are meant to inform stakeholders involved in negotiations of acceptance testing clauses about the implications of including guard bands or exceedance probability levels in contractual specifications.

5.2 Assumptions, Limitations and Recommendations

A number of assumptions and limitations of scope were made in this research, the most prominent of which is the restriction to the use of pyranometer measurements in performance ratio calculations despite the many other use-cases and performance metrics. In this section, the motivation for this and other limitations and assumptions will be outlined, and recommendations for future research are presented.

In literature, energy yield forecasting in the long- and short-term is described as another important use of solar irradiance measurements in utility-scale PV upon which significant financial interests are dependent. Pre-construction yield assessments determine the projected cash flow available for debt-service, which in turn determines the amount of debt the project can take on. Short-term forecasts (intra-day and day-ahead) help the plant operator schedule power output and potentially optimize energy trading strategy. This work does not consider this use-case of pyranometer data, because the relation between ground-measurements and the model output is opaque and method-dependent. Furthermore, modern (short-term) forecasting methods are reliant upon machine-learning methods that are ill-suited for GUM uncertainty propagation due to their highly nonlinear nature. Furthermore, the range of methodologies is larger and less standardized, with some methods exclusively relying upon satellite data. Therefore, it was chosen to limit this study to PV performance metrics. The publication by Reise et al. (2018) is recommended as a starting point for future work investigating uncertainty in this field.

Multiple PV system performance metrics were found to exist in literature, with some having been codified in official standards. The IEC 61724 standard defines the performance ratio and energy performance index, while the ASTM E2848 standard defines an assessment method involving linear extrapolation to reference conditions. This work exclusively considered the performance ratio, due to the other two performance metrics being defined with respect to the results of a computational model of the power plant. Assessing the uncertainty in these performance metrics at the same degree of rigor as done with the PR would require uncertainty propagation through such a computational model, potentially through Monte Carlo methods if not possible analytically. It would moreover introduce a new degree of freedom in the choice of computational model. To maintain a focused scope, it was decided to restrict the analysis of this work to the performance ratio.

From the interviews it was gathered that the ASTM E2848 standard is the preferred PV performance metric in the US utility-PV sector, while the PR is preferred in the rest of the world. For the US-market, a separate calculation quantifying the achievable uncertainty improvements for the ASTM method may be merited. However, the ASTM E2848 method employs a linear extrapolation of the measured data to certain reference conditions, which is a questionable assumption. The epistemic uncertainty of this extrapolation may outweigh the measurement uncertainty. Thus, it is recommended that a potential extension of this work to the ASTM method starts by investigating the validity of this extrapolation.

Further research into uncertainty of computational PV models is recommended. A good starting point for such a study would be the work by Deville et al. (2026), which features uncertainty and bias comparisons for common software packages.

The analysis of the utility-scale solar PV sector consisted of a literature review and a small qualitative component that was restricted to 12 informal conversations and 6 formal interviews with industry experts. While the interviews helped corroborate and add to the findings of the literature study, the generalizability of the interview findings is highly limited due to the small sample size in comparison to the size of the solar PV industry. As such, the sector analysis is mostly based on published information and secondary sources such as best-practice guides. A large-scale review of industry contractual practices based on primary sources such as interviews and contracts would benefit this study by substantiating or further informing the acceptance testing models described in this work.

Furthermore, this work was initially focused on post construction acceptance testing in ECP contracts due to the larger financial interests dependent on these tests. The interviews therefore primarily discussed performance guarantees in EPC contracts and not in O&M contracts. Extension of the sector analysis to include discussions with O&M providers about performance guarantee practices in the O&M market would facilitate an accurate estimation of the applicability of the results to the O&M market.

Finally, interviews revealed that an analysis of uncertainty in performance metric measurement is usually deferred to an independent engineer. Extending the interviews with discussions on the methodology employed in these assessments with these engineers would allow for an interesting comparison to the methodology and assumptions used in this work.

The PR uncertainty calculations relied upon several assumptions, such as normality, linear sensitivity, no cross-correlation and complete autocorrelation. some of which have already been discussed in the previous section. Strong cases can be made that these assumptions were incorrect for some uncertainties. However, given the lack of studies characterizing these covariances, a consistent approach was preferred over a source-by-source individual treatment based on assumptions by the author.

The present assumptions helped identify the directional response as the most significant source of uncertainty in present state-of-the-art pyranometers. However, the applicability of the sensitivity analysis of the total uncertainty with respect to other pyranometer sources should be considered minimal, because the magnitudes of these uncertainties are very similar. Once the directional response error has been reduced, a characterization of the covariances would be recommended to inform subsequent directions of improvement.

Real measurement data was required to properly combine relatively and absolutely defined uncertainty sources. Data from two research-scale PV systems was used for this purpose, which is quite limited. Furthermore, the work aims to advise the utility-scale PV industry. Using data from utility-scale plants would therefore be preferable, but this data is often proprietary and not publicly shared. The calculations required datasets containing timestamped power output, solar irradiance and other environmental data, as well as metadata like the plant location, module power rating and tilt angles. Moreover, this work selected for powerplants with homogeneous module orientation. The findings of this work would be supported by performing calculations with more datasets that span a greater amount of time, to limit the influence of dataset-specific anomalies or fluctuations. Comparison of the results of this work to calculations with data of larger-scale plants would also be a meaningful extension of this work. A potential starting point would be the database of PV plant data used in the report by French et al. (2021), that includes datasets for systems with a capacity of up to 500 kW_p.

There are numerous factors influencing plant performance and uncertainty that were not included in this work, with two prominent examples being degradation and soiling. With regards to factors affecting plant power output, this study is rooted in the philosophy that physically controllable factors should not be corrected for in the PR as their influence is determined through maintenance or construction quality. If the power output is lower due to high degradation or soiling rates, then this

should be reflected through a lower PR. While not present in the data used in this study, this philosophy extends to losses due to inverter clipping, since installing inverter capacity below the plant power rating is a controllable design choice.

However, exclusion of soiling or degradation losses does not apply to irradiance data, since the PR must be an honest comparison between plant output and available solar energy. The philosophy of this work therefore dictates that pyranometer non-stability and soiling should be included. Calculations included non-stability, but not soiling. Soiling rates are highly dependent on local environmental factors, and studies were found to report a wide range of potential soiling rates. It was elected to exclude pyranometer measurement uncertainty due to soiling since this merits its own comprehensive study. Applying an estimated soiling rate and uncertainty in this study would obfuscate the generalizability of the results. The python code in appendix C allows for easy inclusion of a new uncertainty source such as soiling. Future work could expand upon this research by using the code to perform a methodological assessment of the impact of different pyranometer soiling rates and cleaning schedules on the total PR uncertainty.

The calculated PR uncertainty and the achievable improvements in accuracy were placed in the context of performance guarantee conformity assessment, which was identified as an important use case of performance metrics such as the PR. Still, guarantees with respect to PV performance metrics and liquidated damages in case of underperformance are usually only of significance in turnkey project structures, where the EPC contractor and final asset owner are two different entities. Conversely, literature described an aversion to performance guarantees by third-party O&M providers, who are usually unwilling to assume responsibility for a metric largely defined by construction quality. The added value of the financial risk calculations in this work is therefore mostly limited to turnkey projects and O&M agreements in which performance guarantees are provided.

As discussed in sections 2.1.2 and 2.4.2, literature indicates that turnkey project structures are common in the utility-scale PV industry, indicating the applicability of this work is not restricted to a niche segment of the industry. Although it is similarly stated that outsourcing O&M to third-parties is also common, literature also notes an increasing reluctance of third-party O&M companies to provide performance guarantees. Notably no quantifications accompany these claims, further underlining the benefits that an expanded sector analysis into the composition of the PV industry would have to this work.

The financial risk calculations themselves are valid for linear payment structures, but rely upon a certain ‘true PR’ of the power plant, which would be the PR measured by perfect measurement instruments. However, this true PR is itself a quantity dependent on specific environmental conditions during the testing period (e.g. partial shading of the power plant would influence the true PR), making it a somewhat ambiguous quantity. It is difficult to make an a priori assessment of financial risk based on measurement uncertainty when the measurand itself is uncertain without a clear quantification. This unpredictable nature is an inherent limitation of the PR and other metrics, and some level of variability will always persist. The fixing of this ‘true PR’ and the consequences on the wider applicability of the risk assessment should therefore be treated with some caution.

6. Conclusion

This work identified a knowledge gap in the composition of uncertainty in PV performance metrics as a combination of different uncertainty sources, and in the way that uncertainty impacts the application of such metrics throughout the utility-scale solar PV industry. Although significant financial interests can be dependent on performance ratio measurements, little research was found to have been done on how measurement uncertainty translates to financial uncertainty in important uses of the performance ratio, or on how impactful improvements of pyranometers can be in this regard. The following research question was posited in order to fill this gap:

RQ: *What is the influence of individual uncertainty sources in state-of-the-art pyranometers to the uncertainty of utility-scale solar PV plant performance metrics, specifically the performance ratio, and what are the most impactful avenues for pyranometer improvement?*

A multi-methodological approach was used in order to answer this research question, that consisted of three main components: an analysis of the utility-scale PV sector through literature research in combination with a small set of expert interviews, rigorous performance ratio uncertainty calculations following GUM methodology, and an assessment of the financial risk due to measurement uncertainty in an important use-case of performance ratios: conformity assessment to performance guarantees in EPC and O&M contracts.

Uncertainty calculations were performed using uncertainty bounds taken from the standard specifications for ISO 9060 class A pyranometers, and the manufacturer specifications of the Hukx SR300-D1 class A pyranometer. Calculations obtained uncertainty in the performance ratio between 2.8 and 3.9%, which is in line with previously published estimates. The SR300-D1 was found to obtain 0.6% lower PR uncertainty compared to the standard ISO 9060 class A specifications, demonstrating that the current state-of-the-art significantly outperforms the standardized definition of high accuracy. Pyranometer uncertainty was found to be responsible for 75-85% of the total uncertainty in the performance ratio, with the PV module calibration being the second most significant source. The directional response was found to be the primary source of uncertainty in all calculations, contributing to 35-60% the total. The directional response error should therefore be the primary focus of efforts to reduce pyranometer uncertainty. In the ISO 9060 class A specifications, the pyranometer temperature response and pyranometer zero-offset A were also found to be significant contributors to total uncertainty.

Some assumptions were made in these calculations, notably independence of uncertainty sources and full autocorrelation over time. To inform further optimal improvement avenues for pyranometers, specifically for the SR300, it is recommended to perform calculations with covariances that more closely follow physical reality, since no other optimal improvement avenues for the SR300-D1 were identifiable under the present assumptions.

The sector analysis revealed a variety of use-cases of solar irradiance measurements, that were broadly categorized between performance assessment (the focus of this work) and yield forecasting. Multiple widely-used performance metrics are available in the IEC 61724 and ASTM E2848 standards, but this study is restricted to the (temperature-corrected) performance ratio defined in the IEC 6172 standard due to its suitability for uncertainty propagation and independence from computational models. Important applications of these metrics are to inform operational and maintenance decisions, calculate the financial value of the plant, monitor overall plant degradation and assess service quality of construction and maintenance work performed by EPC and O&M contractors. Performance guarantees were found to be common in EPC and O&M contracts, with liquidated damages in case of underperformance being proportional to the percentage underperformance. However, no previous studies investigated how measurement uncertainty affects performance guarantee conformity assessment accuracy and its associated financial risk for the parties to such tests.

Two different conformity assessment tests described in literature were analyzed: simple (guard-banded) comparison, and a test based on exceedance probability levels. In-depth results can be found in section 4.3. According to literature and interviews, simple guard banded comparison is common industry practice.

Calculations showed that the statistically expected liquidated damages increase with the measurement uncertainty, meaning that financial risk for EPC and O&M contractors increases with the uncertainty at the benefit of the asset owner. However, though omnipresent, this effect is overshadowed when the test includes guard bands or exceedance probability levels. These shift the risk of erroneous acceptance or rejection very significantly to one party. Importantly, guard bands shift financial risk to the asset owner, leading to systematic undercompensation to the asset owner for underperforming power plants. At higher measurement uncertainties, this undercompensation is only partially offset by the aforementioned statistical advantage of the asset owner. A more comprehensive analysis of the risk profiles and their fairness is provided in section 4.3. The results of these calculations can be used by asset owners, EPC and O&M contractors to interpret the fairness and risk-shift in common acceptance testing clauses.

In industry-standard performance guarantee tests, expected liquidated damages payable by EPC or O&M contractors were found to be dependent on uncertainty only when the margins of under- or overperformance are of similar magnitude as the measurement uncertainty. In this case, a reduction of uncertainty in PR achieved by the SR300-D1 over standard class A specifications corresponds to a 13% reduction in expected damages. This can be increased to 35% if the directional response error of the SR300-D1 were to be halved compared to present specifications. The value-at-risk was found to be linearly proportional to both the measurement uncertainty and the percentage underperformance, and not restricted to the aforementioned regime. Higher accuracy was therefore seen to provide a considerable reduction of the value-at-risk, with relatively greater impact for better-performing plants. The results demonstrate that pyranometer accuracy improvements can greatly reduce the financial risk for EPC and O&M contractors in post-construction acceptance testing and maintenance quality assessment. Similarly, smaller guard bands can reduce financial risk for asset owners. Therefore, this work advocates narrowing guard bands and investing in higher-accuracy measurement equipment, to ensure fairness of performance guarantee conformity tests.

The results obtained in this research add to the literature by providing new insights in the composition of the uncertainty in the PR and by explicitly calculating the impact of PR uncertainty in acceptance testing, allowing to assess the impact of accuracy improvements in this context. The implementation of contractual performance guarantees is not industry-wide, nor is the usage of PR as the specific performance metric. However, the subclass of the industry to which the financial calculations are applicable remains sizeable. Additionally, the results of the performance metric uncertainty calculations, specifically the uncertainty decomposition and the identification of directional response as most significant error source, are universally applicable.

In order to guide pyranometer improvements beyond the presently suggested directional response error reduction, a more realistic treatment of the involved covariances is required. It is advised to base these recommendations on calculations using datasets with a greater timespan and that concern utility-scale power plants, such that their effect is quantified with greater realism. The code developed for this research can be used to this end without requiring significant extension, and can be found in appendix C. Moreover, inclusion of pyranometer soiling and the associated uncertainty in these calculations is recommended, since this is an important factor excluded from this study.

Finally, an expansion upon the sector analysis and qualitative component of this research would support the development of a more comprehensive understanding of both common and niche industry practices. An extensive sector analysis based on primary sources would be valuable to the utility-scale PV industry. The methods and findings of this study could then be applied to the identified industry practices in performance assessment and acceptance testing, to inform further recommendations to the sector.

Bibliography

AlFaraj, J., Popovici, E., & Leahy, P. (2024). Solar Irradiance Database Comparison for PV System Design: A Case study. *Sustainability*, *16*(15), 6436. https://doi.org/10.3390/su16156436

American Society for Testing and Materials. (2015). *ASTM G183-15: Standard practice for field use of pyranometers, pyrheliometers and UV radiometers* (Vol. 14.04). https://doi.org/10.1520/G0183-15

American Society for Testing and Materials. (2017). *Standard Guide for Evaluating Uncertainty in Calibration and Field Measurements of Broadband Irradiance with Pyranometers and Pyrheliometers* (ASTM Standard No. G213-17). https://store.astm.org/g0213-17.html

American Society for Testing and Materials. (2018). *Standard Test Method for Reporting Photovoltaic Non-Concentrator System Performance* (ASTM Standard No. E2848-13:2018). https://store.astm.org/e2848-13r18.html

Auslender, D., & Emami, B. (2025). Calculation of expanded uncertainty for a capacity test. *Proceedings presented at PVPMC 2025 Workshop*. https://www.sandia.gov/app/uploads/sites/243/dlm_uploads/2025/05/37_Calculation-of-Expanded-Uncertainty-for-a-Capacity-Test.pdf

Bachour, D., Perez-Astudillo, D., & Martin-Pomares, L. (2016). Study of soiling on pyranometers in desert conditions. In *Conference Proceedings EuroSun 2016* (pp. 1–9). https://doi.org/10.18086/eurosun.2016.09.08

Bamisile, O., Acen, C., Cai, D., Huang, Q., & Staffell, I. (2024). The environmental factors affecting solar photovoltaic output. *Renewable and Sustainable Energy Reviews*, *208*, 115073. https://doi.org/10.1016/j.rser.2024.115073

Baringa Partners LLP. (2022). *Commercial power purchase agreements: A market study including an assessment of potential financial instruments to support renewable energy commercial power purchase agreements*. https://advisory.eib.org/publications/attachments/commercial-power-purchase-agreements.pdf

Basson, H., & Pretorius, J. (2016). Risk mitigation of performance ratio guarantees in commercial photovoltaic systems. *Renewable Energy and Power Quality Journal*, *14*(1). https://doi.org/10.24084/repqj14.244

Campos, R. A., Martins, G. L., & Rüther, R. (2022). Assessing the influence of solar forecast accuracy on the revenue optimization of photovoltaic + battery power plants in day-ahead energy markets. *Journal of Energy Storage*, *48*, 104093. https://doi.org/10.1016/j.est.2022.104093

Cebecauer, T., & Suri, M. (2016). Site-adaptation of Satellite-based DNI and GHI Time Series: Overview and SolarGIS Approach. *AIP Conference Proceedings, 1734*, 150002. https://doi.org/10.1063/1.4949234

Curtis, T., Heath, G., Walker, A., Desai, J., Settle, E., & Barbosa, C. (2021). *Best practices at the end of the photovoltaic system performance period* (NREL/TP-5D00-78678). National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy21osti/78678.pdf

Deville, L., Anderson, K. S., Sutterlueti, J., Chambers, T. L., De Brabandere, K., Cicala, F. P., Lopez-Lorente, J., Mirletz, B., Neubert, A., Nikam, M., Oliosi, M., Prilliman, M., Rhee, K., Schnierer, B., Spokes, J., Wittmer, B., & Theristis, M. (2025). Feature review of photovoltaic modeling software utilizing blind performance assessment. *Solar Energy*, *304*, 114207. https://doi.org/10.1016/j.solener.2025.114207

Di Leo, P., Ciocia, A., Malgaroli, G., & Spertino, F. (2025). Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive review. *Energies*, *18*(8), 2108. https://doi.org/10.3390/en18082108

Dierauf, T., Growitz, A., Kurtz, S., Becerra Cruz, J. L., Riley, E., & Hansen, C. (2013). *Weather-corrected performance ratio* (NREL/TP-5200-57991). National Renewable Energy Laboratory. https://docs.nrel.gov/docs/fy13osti/57991.pdf

Dirnberger, D., & Kraling, U. (2013). Uncertainty in PV Module Measurement—Part I: Calibration of Crystalline and Thin-Film modules. *IEEE Journal of Photovoltaics*, *3*(3), 1016–1026. https://doi.org/10.1109/jphotov.2013.2260595

French, R. H., Bruckman, L. S., Moser, D., Lindig, EURAC., van Iseghem, M., ISI, F., Stein, J. S., Richter, M., Herz, M., van Sark, W., & Baumgartner, F. (2021). *Assessment of Performance Loss Rate of PV Power Systems*. International Energy Agency Photovoltaic Power Systems Programme. https://iea-pvps.org/wp-content/uploads/2021/04/IEA-PVPS-T13-22_2021-Assessment-of-Performance-Loss-Rate-of-PV-Power-Systems-report.pdf

Fuke, P., & Kottantharayil, A. (2025). Effect of pyranometer soiling on the PV performance ratio evaluation and correction methods. *Solar Energy*, *298,* 113634. https://doi.org/10.1016/j.solener.2025.113634

Gibbs, B. P. (2011). *Advanced Kalman filtering, Least-Squares and modeling*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470890042

Habte, A., Gueymard, C., Wilbert, S., Lorenz, E., Balenzategui Manzanares, A., Sengupta, M., Myers, D., Stoffel, T., Vignola, F., Yang, J., Lauret, P., & David, M. (2024). *Principles and practical methods for estimating uncertainty and evaluating solar irradiance data*. National Renewable Energy Laboratory. In *Best practices handbook for the collection and use of solar resource data for solar energy applications* (4th ed.). https://doi.org/10.2172/2448063

Hansen, C., & Martin, C. (2015). *Photovoltaic system modeling: Uncertainty and sensitivity analyses* (SAND2015-6700). Sandia National Laboratories. https://doi.org/10.13140/RG.2.1.4444.2081

Hukx. (n.d.-a). *SR300-D1 pyranometer datasheet*. Hukx. Retrieved December 24, 2025, from https://www.hukseflux.com/products/pyranometers-solar-radiation-sensors/pyranometers/sr300-d1-pyranometer

Hukx. (n.d.-b). *Hukx user manual for industrial series Class A and Class B pyranometers*. Hukx. Retrieved January 21, 2025, from https://www.hukx.com/products/sr300-d1-pyranometer

International Electrotechnical Commission. (2020). *Electricity metering equipment — Particular requirements — Part 22: Static meters for AC active energy (classes 0,1S, 0,2S and 0,5S)* (IEC Standard No. 62053-22:2020). https://webstore.iec.ch/en/publication/29987

International Electrotechnical Commission. (2021). *Photovoltaic system performance* (IEC Standard No. 61724:2021). https://webstore.iec.ch/en/publication/65561

International Electrotechnical Commission. (2023). *IEC 60904-2: Photovoltaic devices — Part 2: Requirements for photovoltaic reference devices* (IEC Standard No. 60904-2:2023). https://webstore.iec.ch/en/publication/68536

International Energy Agency. (2024a). *Renewables 2024: Analysis and forecasts to 2030*. https://www.iea.org/reports/renewables-2024

International Energy Agency (2024b), *World Energy Investment 2024*. https://www.iea.org/reports/world-energy-investment-2024

International Energy Agency Photovoltaic Power Systems Programme. (2022). *Guidelines for operation and maintenance of photovoltaic power plants in different climates* (IEA-PVPS T13-25:2022; ISBN 978-3-907281-13-0). International Energy Agency Photovoltaic Power Systems Programme. https://iea-pvps.org/wp-content/uploads/2022/11/IEA-PVPS-Report-T13-25-2022-OandM-Guidelines.pdf

International Finance Corporation. (2015). *Utility-Scale Solar Photovoltaic Power Plants: A project Developer's guide*. https://doi.org/10.1596/22797

International Organization for Standardization. (2018). *Solar energy: Specification and classification of instruments for measuring hemispherical solar and direct solar radiation* (ISO Standard No. 9060:2018). https://www.iso.org/standard/67464.html

International Organization for Standardization. (2023). *Solar energy: Calibration of pyranometers by comparison to a reference pyranometer* (ISO Standard No. 9847:2023). https://www.iso.org/standard/78800.html

Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data – Guide to the Expression of Uncertainty in Measurement.* (JCGM Standard No. 100:2008). https://doi.org/10.59161/JCGM100-2008E

Joint Committee for Guides in Metrology. (2011). *Evaluation of measurement data – Supplement 2 to the "Guide of uncertainty in measurement" – extension to any number of output quantities.* (JCGM Standard No. 102:2011). https://doi.org/10.59161/JCGM102-2011

Joint Committee for Guides in Metrology. (2012). *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment* (JCGM Standard No. 106:2012). https://doi.org/10.59161/JCGM106-2012

Konings, J., & Habte, A. (2015). *Uncertainty evaluation of measurements with pyranometers and pyrheliometers.* In *Proceedings of the Solar World Congress 2015*. International Solar Energy Society. https://doi.org/10.18086/swc.2015.07.15

Leva, S., Nespoli, A., Pretto, S., Mussetta, M., & Ogliari, E. G. C. (2020). PV plant power nowcasting: A real case comparative study with an open access dataset. *IEEE Access*, *8*, 194428–194440. https://doi.org/10.1109/access.2020.3031439

Masson, G., Van Rechem, A., De L'Epine, M., & Jäger-Waldau, A. (2025). *Snapshot of global PV markets 2025*. International Energy Agency Photovoltaic Power Systems Programme. https://doi.org/10.69766/pbhv9141

Meydbray, J., Emery, K., & Kurtz, S. (2012). *Pyranometers and reference cells: What's the difference?* (NREL/JA-5200-54498). National Renewable Energy Laboratory. Published in *PV Magazine*, 04/2012. https://docs.nrel.gov/docs/fy12osti/54498.pdf

Mittler, C., Bucksteeg, M., & Staudt, P. (2025). Review and morphological analysis of renewable power purchasing agreement types. *Renewable and Sustainable Energy Reviews*, *211*, 115293. https://doi.org/10.1016/j.rser.2024.115293

Mohamadi, F. (2021). *Introduction to project finance in renewable energy infrastructure*. https://doi.org/10.1007/978-3-030-68740-3

National Renewable Energy Laboratory, Sandia National Laboratories, SunSpec Alliance, & SunShot National Laboratory Multiyear Partnership (SuNLaMP) PV O&M Best Practices Working Group. (2018). *Best practices for operation and maintenance of photovoltaic and energy storage systems* (3rd ed.; NREL/TP-7A40-73822). National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy19osti/73822.pdf

Newmiller, J., Mikofski, M., & Holmgren, W. F. (2025). Photovoltaic power plant capacity test uncertainty. In *2025 IEEE 53rd Photovoltaic Specialists Conference (PVSC)* (pp. 1347–1352). https://doi.org/10.1109/pvsc59419.2025.11133042

Özkalay, E., Virtuani, A., Fairbrother, A., Skoczek, A., Friesen, G., & Ballif, C. (2022). How does the use of satellite-derived insolation data impact the accuracy of performance ratio estimates?. *Conference Proceeding of the 8th World Conference on Photovoltaic Energy Conversion: European PV Solar Energy Conference (WCPEC: EU PVSEC).* https://aris.supsi.ch/entities/publication/ea7c4837-c57d-48b4-8dd8-835fc13b5760

Pacudan, R. (2015). Financing solar PV projects: Energy production risk reduction and debt capacity improvement. In S. Kimura, Y. Chang, & Y. Li (Eds.), *Financing renewable energy development in East Asia Summit countries* (ERIA Research Project Report No. 2014-27, pp. 297–320). Economic Research Institute for ASEAN and East Asia (ERIA). https://www.eria.org/RPR_FY2014_No.27_Chapter_10.pdf

Polo, J., Wilbert, S., Ruiz-Arias, J., Meyer, R., Gueymard, C., Súri, M., Martín, L., Mieslinger, T., Blanc, P., Grant, I., Boland, J., Ineichen, P., Remund, J., Escobar, R., Troccoli, A., Sengupta, M., Nielsen, K., Renne, D., Geuder, N., & Cebecauer, T. (2016). Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. *Solar Energy*, *132*, 25–37. https://doi.org/10.1016/j.solener.2016.03.001

Prilliman, M. J., Hansen, C. W., Keith, J. M. F., Janzou, S., Theristis, M., Scheiner, A., & Ozakyol, E. (2023). *Quantifying uncertainty in PV energy estimates: Final report* (NREL/TP-7A40-84993). National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy23osti/84993.pdf

PWC. (2024). *Engineering, Procurement and Construction (EPC) Contracts in the solar sector: Investing in energy transition projects.* https://www.pwc.com/m1/en/blogs/pdf/epc-contracts-in-solar-sector.pdf

Raji, L., Razak, A., & Sharol, A. (2025). Comparative assessment of PV simulation tools for a megawatt-scale rooftop solar photovoltaic system in tropical climate. *Next Energy*, *9*, 100452. https://doi.org/10.1016/j.nxener.2025.100452

Reise, C., Müller, B., Moser, D., Belluardo, G., & Ingenhoven, P. (2018). *Uncertainties in PV system yield predictions and assessments* (IEA-PVPS T13-12:2018). International Energy Agency Photovoltaic Power Systems Programme. https://iea-pvps.org/key-topics/uncertainties-in-pv-system-yield-predictions-and-assessments/

Ribeiro, K., Santos, R., Saraiva, E., & Rajagopal, R. (2021). A statistical methodology to estimate soiling losses on photovoltaic solar plants. *Journal of Solar Energy Engineering*, *143*(6). https://doi.org/10.1115/1.4050948

Richter, M., Tjengdrawira, C., Vedde, J., Green, M., Frearson, L., Herteleer, B., Jahn, U., Herz, M., & Köntges, M. (2017). *Technical assumptions used in PV financial models: Review of current practices and recommendations* (IEA-PVPS T13-08:2017). International Energy Agency Photovoltaic Power Systems Programme. https://iea-pvps.org/key-topics/technical-assumptions-used-in-pv-financial-models/

Seel, J., Kemp, J.M., Cheyette, A., Millstein, D., Gorman, W., Jeong, S., Robson, D., Setiawan, R., & Bolinger, M. (2024). *Utility Scale Solar, 2024 Edition: Empirical trends in deployment, technology, cost, performance, PPA pricing, and value in the United States.* Lawrence Berkeley National Laboratory, Energy Markets and Policy Department. https://emp.lbl.gov/publications/utility-scale-solar-2024-edition

Solar Bankability Consortium. (2016). *Review and gap analyses of technical assumptions in PV electricity cost: Report on current practices in how technical assumptions are accounted in PV investment cost calculation* (Deliverable D3.1). https://www.tuv.com/content-media-files/master-content/services/products/p06-solar/solar-downloadpage/solar-bankability_d3.1_review-and-gap-analysis-of-technical-assumptions-in-pv-electricity-cost.pdf

SolarPower Europe. (2021). *Engineering, procurement & construction best practice guidelines: Version 2.0* (ISBN 9789464444209). https://www.solarpowereurope.org/insights/thematic-reports/epc-best-practice-guidelines-version-2-0/

SolarPower Europe (2025). *Operation & Maintenance Best Practice Guidelines: Version 6.0* (ISBN 9789464444247). https://www.solarpowereurope.org/insights/thematic-reports/operation-and-maintenance-best-practice-guidelines-version-6-0-1

Spitters, C., Toussaint, H., & Goudriaan, J. (1986). Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis Part I. Components of incoming radiation. *Agricultural and Forest Meteorology*, *38*(1–3), 217–229. https://doi.org/10.1016/0168-1923(86)90060-2

Stoel Rives LLP. (2022). *The law of solar: A guide to business and legal issues (6th edition)*. https://www.stoel.com/insights/reports/the-law-of-solar/the-law-of-solar

Strobel, M., Betts, T., Friesen, G., Beyer, H., & Gottschalg, R. (2009). Uncertainty in Photovoltaic performance parameters – dependence on location and material. *Solar Energy Materials and Solar Cells*, *93*(6–7), 1124–1128. https://doi.org/10.1016/j.solmat.2009.02.003

Swami, G., Sheth, K., & Patel, D. (2024). PV Capacity Evaluation using ASTM E2848: Techniques for Accuracy and Reliability in Bifacial Systems. *Smart Grid and Renewable Energy*, *15*(09), 201–216. https://doi.org/10.4236/sgre.2024.159012

Villacorta, M. C., & Victoria, M. (2024). Design and operation of utility-scale PV power plants. In *Fundamentals of solar cells and photovoltaic systems engineering* (Chap. 10, pp. 337–364). Elsevier. https://doi.org/10.1016/B978-0-323-96105-9.00010-0

Visser, L., AlSkaif, T., Khurram, A., Kleissl, J., & Van Sark, W. (2024). Probabilistic solar power forecasting: An economic and technical evaluation of an optimal market bidding strategy. *Applied Energy*, *370*, 123573. https://doi.org/10.1016/j.apenergy.2024.123573

Warade, S., & Kottantharayil, A. (2018). Analysis of soiling losses for different cleaning cycles. *Conference proceedings of the 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)* (pp. 3644–3647). IEEE. https://doi.org/10.1109/PVSC.2018.8547867

Wolfe, P.R. (2013). *Defining 'utility-scale' solar: How we arrived at the threshold of 4 MW-AC*. Wiki-Solar Glossary. https://wiki-solar.org/library/papers/WSG02-02_Utility-scale.pdf

Appendices

A: Uncertainty in the GUM Framework

Due to the dense mathematical content, these derivations were written down in latex, inserted below.

A.1 Outline of the Derivation

We propagate our uncertainty according to the GUM uncertainty framework for explicit multivariate measurement models (GUM, 2012). The guide provides a very concise mathematical formulation for the expression of uncertainty. To avoid any ambiguity, this section contains an explicit derivation of the formulae for total uncertainty and time-aggregated uncertainty used in this work, including all assumptions used to arrive at this result.

We start by writing down the general equations as stated in the GUM.

- An explicit multivariate measurement model specifies a relationship between an output quantity $\mathbf{Y} = (Y_1, ..., Y_m)^{\mathrm{T}}$ and an input quantity $\mathbf{X} = (X_1, ..., X_N)^{\mathrm{T}}$, and takes the form

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}), \quad \text{with} \quad \mathbf{f} = (f_1, ..., f_m)^{\mathrm{T}} \tag{A.1}$$

  Note: any particular function f_j is allowed to depend on a subset of $\mathbf{X}$.

- Given an estimate $\mathbf{x}$ of $\mathbf{X}$, an estimate of $\mathbf{Y}$ is $\mathbf{y} = \mathbf{f}(\mathbf{x})$.

- The covariance matrix of dimension $m \times m$ associated with $\mathbf{y}$ is

$$\mathbf{U_y} = \begin{bmatrix} u(y_1, y_1) & \dots & u(y_1, y_m) \\ \vdots & \ddots & \vdots \\ u(y_m, y_1) & \dots & u(y_m, y_m), \end{bmatrix} \tag{A.2}$$

  where $\mathrm{cov}(y_j, y_j) = u^2(y_j)$, and is given by

$$\mathbf{U_y} = \mathbf{C_x} \mathbf{U_x} \mathbf{C_x}^{\mathrm{T}}. \tag{A.3}$$

  $\mathbf{C_x}$ is the sensitivity matrix of dimension $m \times N$ given by evaluating

$$\mathbf{C_X} = \begin{bmatrix} \frac{\partial f_1}{\partial X_1} & \dots & \frac{\partial f_1}{\partial X_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial X_1} & \dots & \frac{\partial f_m}{\partial X_N} \end{bmatrix} \tag{A.4}$$

  at $\mathbf{X} = \mathbf{x}$.

As an example, consider the measurement equation for the pyranometer measurement for the irradiance in a hemisphere:

$$G = \frac{V}{S} = f(V, S), \tag{A.5}$$

where we have uncertainties u_V and u_S in V and S respectively. One might be tempted to immediately apply the described uncertainty framework to this equation, but it is important to proceed with caution here. The intuition to immediately apply the framework is only correct in the specific cases where *the uncertainty in V and S is not the result of more than a single uncertainty source* or *all uncertainty sources acting on a variable are fully correlated with each other.* In case the error in one or both of these quantities is the combined error of multiple uncertainty sources that are not fully cross-correlated, the direct application of the framework will yield an erroneous result when calculating time-aggregates of G.

The ASTM G213-17 excel spreadsheet for pyranometer uncertainty calculation indirectly employs the GUM methodology, which is the reason such explicit attention is given to this matter. In the next section, we will illustrate where and why the method of the ASTM spreadsheet fails. After that, we will demonstrate the mathematically correct calculation.

A.2 Erroneous Calculation of the Uncertainty

Let us reproduce the error made in the ASTM G213-17 spreadsheet. Suppose we have a time-series of irradiance measurements G_t which we calculate through time-series V_t and S_t. For the sake of simplicity, let us assume there is only an uncertainty in V_t. We assume the uncertainties $u_{V,t}$ in V_t are due to sets of *independent* uncertainty sources, each their own timeseries $u_{V_i,t}$,

$$u_{V,t} = \sqrt{\sum_i u_{V_i,t}^2}. \tag{A.6}$$

The uncertainty in G_t is then

$$\begin{aligned} u_{G,t} &= (\frac{\partial G}{\partial V})_t u_{V,t} \\ &= \sqrt{\sum_i \left((\frac{\partial G}{\partial V})_t u_{V_i,t}\right)^2}, \end{aligned} \tag{A.7}$$

as we would expect for independent uncertainty sources. So far, the method of the ASTM spreadsheet is mathematically correct. In the correct GUM methodology we continue using the lower equality, keeping all individual uncertainty sources explicit. The ASTM method works using the upper equality, it aggregates the individual sources to a total uncertainty in V_t, which will lead to problems in the uncertainty calculation for time aggregates of G.

While we assumed that the uncertainty sources were mutually independent from each other (i.e. no cross-correlation), for time-aggregation we need to include temporal autocorrelation for each independent source. This is where it gets problematic for the ASTM spreadsheet, since there is no clear way to obtain an autocorrelation for an aggregate uncertainty of sources that each have their own autocorrelation coefficients.

Regardless of the method by which these 'total autocorrelation coefficients' $c_{V;t,\tau}$ are determined, we will now finish by calculating the uncertainty of the total solar irradiation $G = \sum_t G_t$. The uncertainty in G is given by

$$\begin{aligned} u_G &= \sum_{t,\tau} u_{G,t}\, u_{G,\tau}\, c_{G;t,\tau} \\ &= \sum_{t,\tau} (\frac{\partial G}{\partial V})_t u_{V,t}\, (\frac{\partial G}{\partial V})_\tau u_{V,\tau}\, c_{V;t,\tau} \\ &= \sum_{t,\tau} \sqrt{\sum_i \left((\frac{\partial G}{\partial V})_t u_{V_i,t}\right)^2} \sqrt{\sum_j \left((\frac{\partial G}{\partial V})_\tau u_{V_j,\tau}\right)^2}\, c_{V;t,\tau} \\ &= \sum_{t,\tau} (\frac{\partial G}{\partial V})_t (\frac{\partial G}{\partial V})_\tau \sqrt{\sum_{i,j} \left(u_{V_i,t}\right)^2 \left(u_{V_j,\tau}\right)^2}\, c_{V;t,\tau} \\ &= \sum_t u_{G,t} + \sum_{t\neq\tau} (\frac{\partial G}{\partial V})_t (\frac{\partial G}{\partial V})_\tau \sqrt{\sum_{i,j} \left(u_{V_i,t}\right)^2 \left(u_{V_j,\tau}\right)^2}\, c_{V;t,\tau} \end{aligned} \tag{A.8}$$

In the last equality, we used the requirement that uncertainty is always fully autocorrelated with itself on the same timestep ($c_{V;t,t} = 1$). In case any correlation coefficient $c_{V;t,\tau}$ is nonzero, the uncertainty in G will include *cross-terms between uncertainty sources*. However, we started with the assumption that all sources of uncertainty were mutually independent, thus we have found a contradiction.

A.3 Correct Calculation of the Uncertainty

We have seen how naive application of the uncertainty framework leads to erroneous results, such as unclear correlations and cross-source contributions where sources ought to be independent. These errors resulted from an implicit mistake we made by grouping the various uncertainties in the quantities V and S together at the start.

The GUM uncertainty framework assumes that a given input quantity has just a single total uncertainty that can have only a single correlation over time or with other sources. Because we started by aggregating our various independent uncertainty sources to a single uncertainty per quantity, it was by construction unavoidable that we would eventually encounter cross terms between independent sources acting on the same quantity.

We can avoid this mistake by thinking about how an instantaneous value of a quantity X_i with K_i different uncertainty sources is defined.

To start, we restrict ourselves to the case of a single output quantity $Y = f(X_1, ..., X_N)$, i.e. $m = 1$ in equation A.1. Say we measure the value of a quantity X_i to be $\tilde{X}_i$ and all uncertainties assume instantaneous values $\tilde{q}_{X_i,k}$, then the actual value $\tilde{x}_i$ must be

$$\tilde{x}_i = \tilde{X}_i - \sum_k \tilde{s}_{i,k}\tilde{q}_{X_i,k}, \tag{A.9}$$

where the $\tilde{s}_{i,k}$ denotes the sensitivity of quantity X_i to uncertainty source k.

Observe that we can rewrite equation A.9 to an equation for quantity X_i:

$$\begin{aligned} X_i &= x_i + \sum s_{i,k} q_{X_i,k} && \text{(A.10)} \\ &= g(x_i,\, \{q_{X_i,k}\}) && \text{(A.11)} \end{aligned}$$

in which x_i is the variable we are trying to measure and $q_{X_i,k}$ are the random variables associated with the uncertainty sources acting on quantity X_i. The sensitivities are defined as

$$s_{i,k} = \frac{\partial g}{\partial q_{X_i,k}} = \begin{cases} 1, & \text{if source k is an absolute error,} \\ x_i, & \text{if source k is a relative error.} \end{cases} \tag{A.12}$$

We see that our quantity X_i is not a basic variable but is itself a measurement equation of the quantities $(x_i,\, \{q_{X_i,k}\})$. Thus, for proper evaluation of the uncertainty in $Y = f(X_1, \ldots, X_N)$ we should explicitly include these dependencies, which is what we neglected to do in section A.2. Let us press on and see what the consequences will be if we keep the dependencies of X_i explicit in our derivation. We will simplify the subscript X_i to simply i from here on.

Before we start building the covariance matrix for the variables $(x_i,\, \{q_{i,k}\})$, observe that we can disregard the quantity x_i from our calculations, due to the fact that x_i is not a random variable but instead the quantity we are trying to measure. Because it is not a random variable it will have no uncertainty nor be covariant with any of the uncertainties $\{q_{i,k}\}$. Mathematically, this means that its associated subspace in the covariance matrix will be entirely 0, and hence we exclude it for clarity.

We will now assume, as we did in section ??, that all uncertainty sources are completely independent of each other. A discussion of this assumption is provided in section ??. The correlations $c_{i,k,l}$ will thus be

$$c_{i,k,l} = \begin{cases} 1, & k = l, \\ 0, & k \neq l. \end{cases} \tag{A.13}$$

The matrix elements of the covariance matrix are then

$$u(q_{i,k}, q_{i,l}) = u_{i,k} u_{i,l}\, c_{i,k,l} = u_{i,k}^2 \tag{A.14}$$

and the covariance matrix will assume the trivial form

$$U_{X_i} = \begin{bmatrix} u(q_{i,1}, q_{i,1}) & \cdots & u(q_{i,1}, q_{i,K_i}) \\ \vdots & \ddots & \vdots \\ u(q_{i,K_i}, q_{i,1}) & \cdots & u(q_{i,K_i}, q_{i,K_i}) \end{bmatrix} = \operatorname{diag}\left(u_{i,1}^2,\, \ldots\, , u_{i,K_i}^2\right). \tag{A.15}$$

We can now obtain the instantaneous uncertainty in X_i by using equation A.3 with the sensitivity matrix

$$C_{X_i} = \begin{bmatrix} s_{i,1} & \cdots & s_{i,K_i} \end{bmatrix}, \tag{A.16}$$

which evaluates to the expected result

$$\begin{aligned} u_i^2 = C_{X_i} U_{X_i} C_{X_i}^{\mathrm{T}} &= \sum_{k,l} s_{i,k} u_{i,k}\; s_{i,l} u_{i,l}\; c_{i,k,l} \\ &= \sum_k (s_{i,k} u_{i,k})^2. \end{aligned} \tag{A.17}$$

Moreover, the uncertainty for a quantity $Y = f(X_1, \ldots, X_N)$ with independent input quantities X_i is a similarly straightforward expression. Under the assumption that all uncertainty sources acting on all our input quantities are independent from each other, the correlations $c_{i,j,k,l}$ will be

$$c_{i,j,k,l} = \begin{cases} 1, & i = j \text{ and } k = l, \\ 0, & \text{else.} \end{cases} \tag{A.18}$$

Again, U_X is diagonal with elements u_i^2. The result for the uncertainty in Y is then, using equations A.3 and A.4 with $m = 1$:

$$\begin{aligned} u_Y^2 = \sum_i \left(\frac{\partial f}{\partial X_i} u_i\right)^2 &= \sum_i \sum_k \left(\frac{\partial f}{\partial X_i}\right)^2 \left(\frac{\partial X_i}{\partial q_{i,k}}\right)^2 u_{i,k}^2 \\ &= \sum_{i,\,k} (S_{i,\,k} u_{i,k})^2, \end{aligned} \tag{A.19}$$

where we introduced the notation $S_{i,\,k_i}$ to denote the total sensitivity of variable Y to uncertainty q_{i,k_i} for the sake of brevity. Observe that the expression is nearly identical to equation A.17, except for the presence of a double sum which can be removed by introducing a new index.

The results obtained so far are identical to those in section ??. Where the separate inclusion of each uncertainty source makes a difference is upon generalization of our expression to multiple output quantities, such as a timeseries of Y.

The formal extension of this formalism to a timeseries of output quantities as prescribed in the GUM is obtained by vectorizing Y and f:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} \qquad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_T \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{i,t} \\ \vdots \\ X_{N,T}, \end{bmatrix} \tag{A.20}$$

with $Y_t = f_t(\mathbf{X})$. In the formalism, f_t takes the entire vector X as input but is not necessarily dependent on each element. In our case, f_t is only dependent on the elements of the same timestep: $\{X_{1,t}, \ldots, X_{N,t}\}$, while the measurement equation itself remains the same function f over time. The sensitivities are thus:

$$\frac{\partial f_t}{\partial X_{i,\tau}} = \begin{cases} (\frac{\partial f}{\partial X_i})_t & \text{if } t = \tau \\ 0 & \text{if } t \neq \tau. \end{cases} \tag{A.21}$$

On each variable may act an arbitrary number of uncertainty sources $q_{i,k}$ with time-dependent magnitudes $u_{i,k,t}$. We remain in our assumption that sources are independent, but we will allow uncertainty sources to be autocorrelated over time, leading to correlation coefficients

$$c_{i,j,k,l,t,\tau} = \begin{cases} c_{i,k}(t,\tau), & i = j \text{ and } k = l, \\ 0, & \text{else.} \end{cases} \tag{A.22}$$

Temporal autocorrelation is very common in uncertainties. For uncertainty sources that are fixed but unknown offsets between readings and measured values, the temporal autocorrelation in the uncertainty is 1. If an uncertainty source is completely random over time the autocorrelation is 0. Exponential decay in autocorrelation over time is also common, in which case it is given by $\exp(-\frac{|t-\tau|}{T_c})$, with T_c the correlation time.

With this definition for our correlations we are equipped to write down the full form of the matrix elements of the sensitivity and covariance matrices:

$$\mathbf{C_X}_{\,t,(i,k,\tau)} = \begin{cases} (\frac{\partial f}{\partial X_i})_t \, (s_{i,k})_t \,, & t = \tau, \\ 0, & \text{else,} \end{cases} \tag{A.23}$$

$$\mathbf{U_X}_{(i,k,t),(j,l,\tau)} = \begin{cases} (u_{i,k})_t (u_{i,k})_\tau c_{i,k}(t,\tau) \,, & i = j \text{ and } k = l, \\ 0, & \text{else.} \end{cases} \tag{A.24}$$

Thus, again using equation A.3, we obtain as covariance matrix elements for our timeseries $\mathbf{Y}$:

$$\mathbf{U}_{\mathbf{Y}\,t,\tau} = \sum_{i,k} \Big(\frac{\partial f}{\partial X_i}\, s_{i,k} u_{i,k}\Big)_t \Big(\frac{\partial f}{\partial X_i}\, s_{i,k} u_{i,k}\Big)_\tau c_{i,k}(t,\tau) \tag{A.25}$$

From this expression we can now evaluate the instantaneous variance of Y by simply setting $t = \tau$

$$u_Y^2(t) = \sum_{i,k} \Big(\frac{\partial f}{\partial X_i}\, s_{i,k} u_{i,k}\Big)_t^2, \tag{A.26}$$

and we can also readily calculate the uncertainty of time-aggregates of the timeseries $\mathbf{Y}$. Time-aggregation is simply an evaluation of the equation

$$Z = \sum_t Y_t, \quad \text{which yields sensitivity vector elements} \quad \mathbf{C}_{\mathbf{Y}\,t} = \frac{\partial Z}{\partial Y_t} = 1. \tag{A.27}$$

Hence, the uncertainty of the time-aggregate is given by

$$u_Z^2 = \mathbf{C}_\mathbf{Y}\mathbf{U}_\mathbf{Y}\mathbf{C}_\mathbf{Y}{}^T = \sum_{t,\tau}\sum_{i,k} \Big(\frac{\partial f}{\partial X_i}\, s_{i,k} u_{i,k}\Big)_t \Big(\frac{\partial f}{\partial X_i}\, s_{i,k} u_{i,k}\Big)_\tau c_{i,k}(t,\tau). \tag{A.28}$$

This is the final expression for the uncertainty of a time-aggregate of a multivariate time-dependent quantity with an arbitrary number of uncertainty sources acting on each dependency.

A.4 Sensitivity to Independent Uncertainty Sources

An advantage of independent uncertainty sources is that it allows for an easy analytical expression for the sensitivity of the total uncertainty to a single uncertainty source. In this research, we are specifically interested in the effect that increasing or decreasing a specific uncertainty bound has on the total uncertainty of the performance ratio.

The uncertainty $u_{i,k}(t)$ due to a specific source may be constant, dependent on the value of the quantity X_i at time t or may even be given by an equation dependent on third variables (as is the case with the directional response error). The uncertainty bounds defined in standards such as the ISO 9060 set a *scale* to the values a specific uncertainty can assume; the uncertainty is separable into a function $\phi_{i,k}(t)$ and a constant scale factor $\sigma_{i,k}$ that is defined by the relevant standard. The expression for the total uncertainty then becomes

$$u_{i,k}(t) = \phi_{i,k}(t)\sigma_{i,k}. \tag{A.29}$$

Substituting expression A.29 into equation A.28 makes the role of the uncertainty bounds explicit,

$$u_Z^2 = \sum_{t,\tau}\sum_{i,k} \Big(\frac{\partial f}{\partial X_i}\, s_{i,k}\phi_{i,k}\Big)_t \Big(\frac{\partial f}{\partial X_i}\, s_{i,k}\phi_{i,k}\Big)_\tau \sigma_{i,k}^2 c_{i,k}(t,\tau) = \sum_{i,k} \beta_{i,k}^2 \sigma_{i,k}^2, \tag{A.30}$$

where we defined

$$\beta_{i,k}^2 = \sum_{t,\tau} \Big(\frac{\partial f}{\partial X_i}\, s_{i,k}\phi_{i,k}\Big)_t \Big(\frac{\partial f}{\partial X_i}\, s_{i,k}\phi_{i,k}\Big)_\tau c_{i,k}(t,\tau). \tag{A.31}$$

Note that simplification A.30 is only possible for independent uncertainty sources. Writing the variance u_Z^2 in this way makes it easy to assess the impact of multiplying the uncertainty bound of source (j,l) by a factor α to the total uncertainty,

$$\sigma_{j,l} \;\to\; \alpha\,\sigma_{j,l} \quad \Rightarrow \quad u_Z^2 \;\to\; \beta_{j,l}^2\alpha^2\sigma_{j,l}^2 + \sum_{\substack{i,k \\ (i,k)\neq(j,l)}} \beta_{i,i}^2\sigma_{i,k}^2. \tag{A.32}$$

Let us conclude this analysis by calculating the partial derivative of the uncertainty u_Z with respect to a bound $\sigma_{j,l}$:

$$\begin{aligned} \frac{\partial u_Z}{\partial \sigma_{j,l}} &= \frac{\partial}{\partial \sigma_{j,l}} \sqrt{\sum_{i,k} \beta_{i,k}^2 \sigma_{i,k}^2} \\ &= \frac{\beta_{j,l}^2 \sigma_{j,l}}{\sqrt{\sum_{i,k} \beta_{i,k}^2 \sigma_{i,k}^2}} = \frac{\beta_{j,l}^2 \sigma_{j,l}}{u_Z}. \end{aligned} \tag{A.33}$$

We see that the sensitivity scales with the relative weight of $\beta_{j,l}^2$ and the magnitude of the bound itself; the uncertainty in Z is more sensitive to the contribution of relatively large uncertainty sources. In other words, the marginal impact to reducing u_Z by reducing $\sigma_{j,l}$ is decreasing.

B: Constructing a Simple Optimal Conformity Test

Recall figures 4.10-4.12, depicting the expected amount of payable damages by EPC & O&M contracting for different acceptance testing clauses, given by equation (3.18). Below, the results for acceptance testing type 1 with $\delta = 0\%$ and of type 2 with $k = -1$ are repeated.

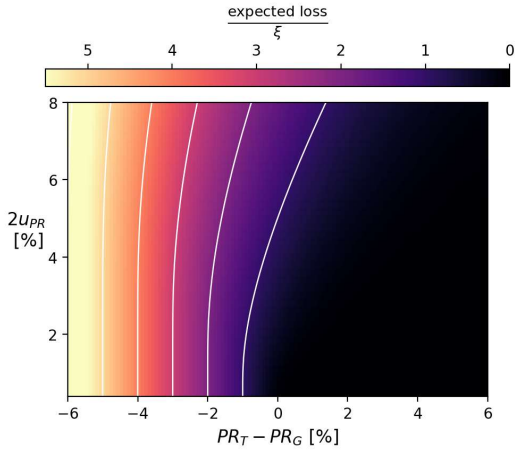


Figure C.1a): Expected payable damages by the contractor for acceptance type 1 and $\delta = 0\%$. White lines denote isoquants.

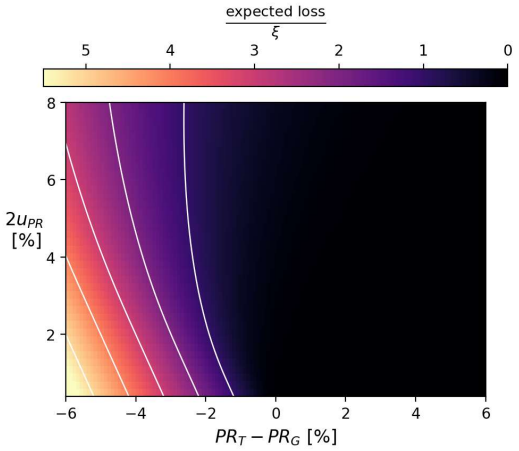


Figure C.1b: Expected payable damages by the contractor for acceptance type 2 and $k = -1$. White lines denote isoquants.

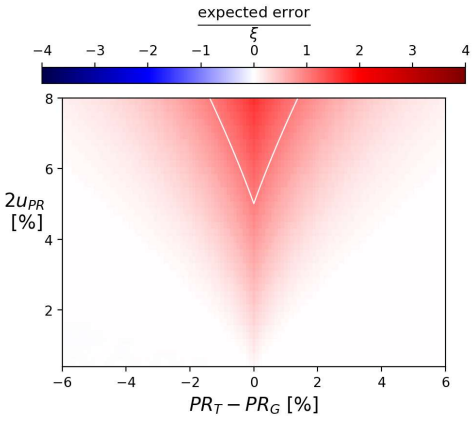


Figure C.1c: Error in the expected payable damages compared to the fair amount for C.1a.

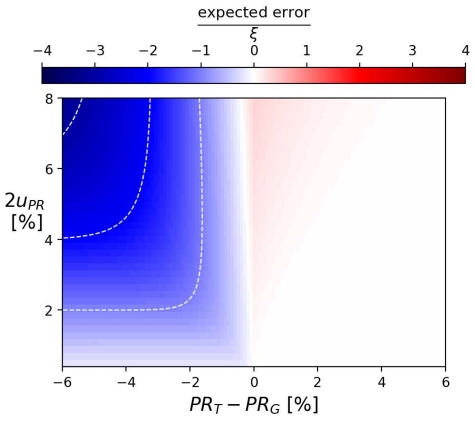


Figure C.1d: Error in the expected payable damages compared to the fair amount for C.1b.

In an ideal situation, the expected owed damages are an exact reimbursement of the discounted value of the missed energy yield. In the above figures, this would be visualized by perfectly vertical isoquants and white error-graphs.

We can attempt to create an 'optimal' acceptance criterion that is a combination of the above two criteria to minimize the error in the statistically expected payable damages. We see that for low u_{PR}, acceptance type 1 exhibits quite vertical behavior, while for higher u_{PR} the acceptance type 2 with negative k exhibits vertical behavior. An acceptance criterion could be constructed that is dependent on the measurement uncertainty could be constructed by following type 1 with $\delta = 0\%$ if the uncertainty is low, and correcting using type 2 in case uncertainty is greater than a certain value. In equation form, this looks like

$$PR_R = PR_G + k \cdot u_{PR}\theta(u_{PR} - m). \qquad (C.1)$$

With θ the Heaviside step function. The factor k determines the strength of the correction term, while the parameter m is a lower cutoff for this term. The parameters k and m can be obtained through a simple optimization using python. Using the

SciPy library, we minimize the loss function, defined as the root-sum-square of the error in the payable damages in the visualized region, using the 'Powell' optimization algorithm.

Initial guesses were taken as $k = -0.5$, $m = 1$. Optimization yielded $k = -0.3187$ and $m = 1.3494$.

The expected payable damages and probability of passing the acceptance test are visualized in figure C.2. In figure C.3, a zoomed-in comparison of the expected error of figure C.1c and the error of this criterion is provided.

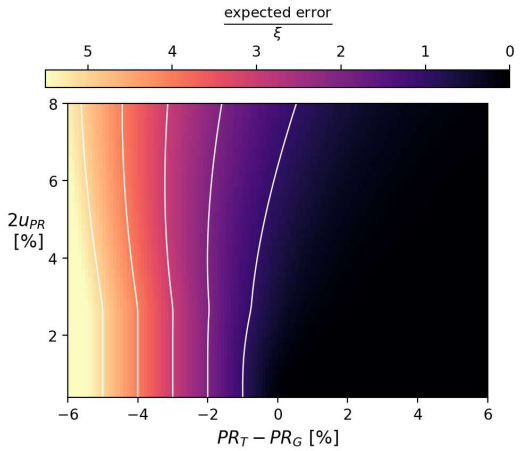


Figure C.2a: Expected payable damages by the contractor for the optimized criterion.

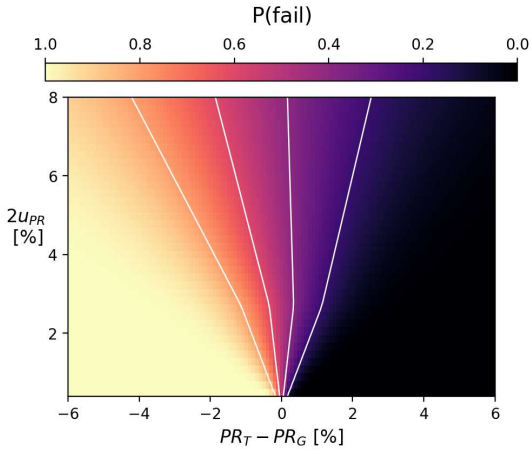


Figure C.2b: Probability of failing the test using the optimized criterion.

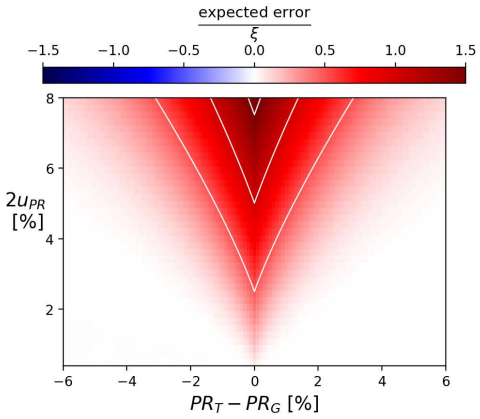


Figure C.3a: Zoomed-in version of figure C.1c.

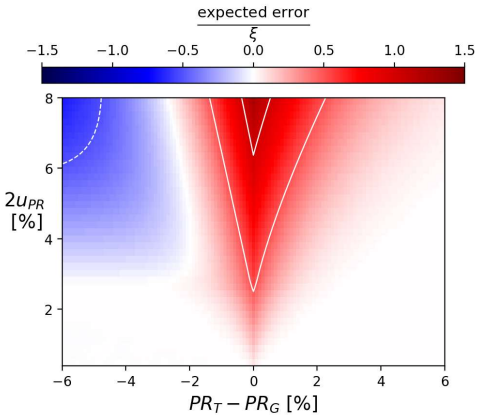


Figure C.3b: Error in the expected payable damages for the optimized criterion.

The criterion described in equation (C.1) requires knowledge of the uncertainty in the PR measurement. We can also try to construct an optimized version of acceptance criterion of type 1, which is not reliant on a measurement uncertainty assessment. Similar optimization as before with an initial guess of $\delta = 0\%$ achieved an optimized value $\delta = 0.2762\%$, the error can be compared for both criteria in figure C.4.

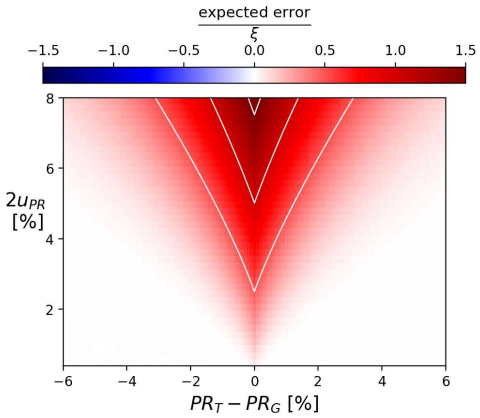


Figure C.4a: Zoomed-in version of figure C.1c.

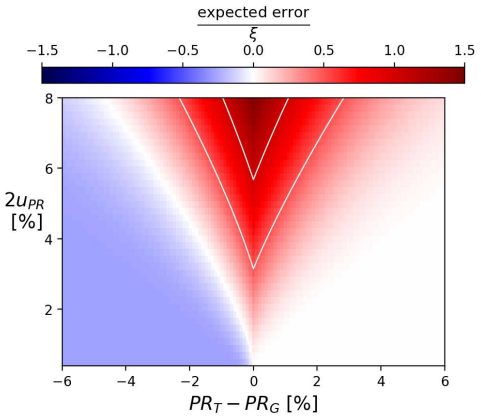


Figure C.4b: Error in the expected payable damages for the optimized criterion.

Importantly, the achieved optimal criteria are determined through several subjective criteria. For instance, the optimization is performed for the grid shown in the figures. However, according to literature and the results of this work, it is very rare for uncertainty to be in the order of 0-2% or >7%. The choice of parameter space that is optimized over influences the result of the optimization. Moreover, no significant study was done with respect to this optimization or investigating whether the achieved values are indeed optima instead of local minima. These results should therefore not be interpreted as recommendations for practical application, but as an illustration to the potential improvement that optimizing a performance criterion can offer.

While optimizations of acceptance criteria to reduce overall error are theoretically interesting, it remains questionable whether industry is willing to adopt such formulations in their contracts. As stated by the interviewed companies, uncertainty analysis is usually not performed by in-house experts but by an independent engineer, if it is analyzed at all. Specifying contractual payments on an optimized criterion that is partially based on the findings of an independent engineer and partially based on an a priori theoretical optimization may lead to highly complicated clauses in EPC and O&M contracts. Thus, while it is theoretically beneficial to optimize acceptance criteria, it may be more practical to stick with industry practice to avoid legal ambiguity.

Furthermore, the limitations on the risk calculations discussed in section 5.2 also apply here: the risk calculations are based on a 'true PR', which remains an ambiguous value that is dependent on the specific environmental conditions during the test.

C: User Guide to the Python Code

The python code developed for this research can be found in the following repository, accompanied by user instructions:

https://github.com/MatthijsRepository/Uncertainty-Propagation